ENZYME-CATALYZED ACYLIUM ION FORMATION AND REAGENTS FOR THE DETECTION OF THIOCARBOXYLATES INVOLVED IN THE BIOSYNTHESIS OF THE NAD DERIVED PINCER COFACTOR

A Dissertation

by

LINA XIE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Tadhg P. Begley
Committee Members,	Frank M. Raushel
	Coran M. H. Watanabe
	Gregory D. Reinhart
Head of Department,	Simon W. North

December 2018

Major Subject: Chemistry

ABSTRACT

In this dissertation, I describe: 1) the first case of experimental trapping of the acylium ion intermediates formed in an enzyme (1-H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase, QDO) catalyzed reaction using interdisciplinary approaches; 2) the computational supports on the QDO acylium ion involved mechanism; 3) the selectively trapping of the protein thiocarboxylates generated from a NAD derived thiocarboxylate cofactor (Nickel pyridinium-3,5-bisthiocarboxylic acid mononucleotide, NiP2TMN) biosynthesis and its partner enzyme (lactate racemase, LarA) activation pathway in the bacterial proteomes using the aromatic sulfonyl azide "click" probes; 4) the selectively detection of the small thiocarboxylate metabolites involved in the NiP2TMN pathway in the bacterial metabolomes using the similar "click" strategies mentioned in 3).

Two acylium ion intermediates in the QDO catalyzed reaction were successfully trapped by the solvent water. The trapping was monitored by ^{18}O isotope labeling assays. The design and synthesis of the substrate analog, 3,7-dihydroxyquinolin-4(1H)-one, and the QDO W153G protein mutant were crucial for the solvent trapping. The former stabilized the first acylium ion intermediate by the enhanced electron donating effect, while the latter opened up the QDO binding pocket for the solvent water better approaching the acylium ion intermediates. DFT energy profile and the MM optimized QDO W153G structure with the substrate docking into the binding pocket provided the computational support for the acylium ion mechanism.

The active form of lactate racemase (LarA) in *Lactobacillus plantarum* had a cofactor NiP2TMN covalently bound to lys184, which was obtained by the coexpression of the complete operon larABCDE. Lissamine Rhodamine B Sulfonyl Azide (LRSA) and biotin-4-

carboxybenzene sulfonyl azide (BiotinSA) were used for selectively trapping the two protein thiocarboxylates, LarA-COSH, LarE-COSH, generated in this NiP2TMN biosynthesis and LarA activation pathway in the bacterial proteomes. BiontinSA labeled LarA-COSH from the protein mixtures was enriched using the streptavidin resin and characterized by LC-MS/MS peptide analysis. The LRSA and dansyl azide (DanA) were further used to specifically trap the two small thiocarboxylate metabolites, pyridinium-3-carboxy-5-thiocarboxylic acid mononucleotide (PCTMN) and pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN) in the bacterial metabolomes. High-resolution LC-MS and LC-MS/MS techniques were used to characterize the "click" labeling products.

循善持真，爱无止息

Love is patient, love is kind

Love rejoices with the truth

Love never fails

--1 Corinthians 13:4, 6 & 8

To my parents, husband, daughter and faith…

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my research advisor, Dr. Tadhg P. Begley, for opening up the door of interdisciplinary research in front of me and guiding me to an astonishing tour on seeking for the new chemistry in lives. Tadhg is such a great scientist, educator, and life mentor. He is always so supportive in both the spirit (knowledge, new ideas, high quality of scientific training, encouragement, patience, and wisdom of facing pressure and failure) and the material (research fundings, lab device purchasing, and sending us to top scientific conferences) aspects. I was amazed by his deep love in science, his art of creative thinking, his passion and ambition on understanding the most complicated reaction mechanisms catalyzed by enzymes, and his open-mind to all the potentially useful approaches. By standing on the shoulder of the giant, I was so excited to see some most beautiful "scenes" that were never discovered by others in the world before! Besides, I also learned how to balance the hard work of scientific research and family life from Tadhg. With his full understanding and blessing, I was able to finish my Ph.D. research smoothly as well as give birth to my healthy and pretty daughter before my defense. I enjoyed the research lifestyle in Tadhg's lab so much, and have made up my mind to continue my academic career on understanding the complicated biosystems in the future.

I express my great thanks to my committee members, Dr. Frank Raushel, Dr. Coran Watanabe and Dr. Gregory Reinhart, for their guidance and support throughout the course of my research. Thank you for providing me with a lot of valuable advice in improving my experimental design and further broadening my view of science.

I also would like to thank all my talented, friendly and supportive colleagues in Begley group. I am so grateful that we could gather together here for the best science from different

corners of the world. I enjoyed all our wonderful discussions in science and beyond. It was a unique experience to work on the interdisciplinary research in the productive multicultural environment. I specially give my deepest gratitude to Hui, BJ, Lisa, Sameh, Dima, Angad, Yiquan, Yindrila, and Jan for the trainings in molecular biology, analytical and synthetic chemistry techniques. Thank Hui for making some important protein mutants, QDO W153G and W153A, for my research in Chapter 2 as the manager of the protein facility. Thank Sameh for providing me with an important synthetic compound, biotin-4-carboxybenzene sulfonyl azide, in Chapter 5 of this research.

I give my special thanks to my collaborator, Dr. Lisa Perez—the manager of the Laboratory for Molecular Simulation at Texas A&M University, for the DFT energy profile calculations on the QDO catalyzed reaction (Chapter 3). I also thank her for training me using the DFT and MM methods for more applications. I also thank the Hausinger group at Michigan State University for providing me with all the important plasmids I needed for my thiocarboxylate labeling proteomic and metabolomic studies (Chapter 5 and 6). I also thank Dr. Jamison P Huddleston from Prof. Frank Raushel's lab for providing me with the trainings on using the EFI-EST-GNT bioinformatics tools.

I also give my thanks to all the on-campus facilities I have used for my research in Texas A&M University, such as the protein chemistry lab in the biochemistry/biophysics department, the mass spectrometry, NMR, and EPR facilities in the chemistry department, and the gene technique lab in the biology department.

I must give my heartfelt thanks to my parents who gave me a healthy life, a brain filled with curiosity to the world and a spirit of never giving up. I cannot say more thanks to my husband, Dapeng, who always takes good care of me and comforts me when I am weak or facing

challenges. I thank the coming of my lovely daughter, Gloria. She makes me stronger, more responsible, and closer to the perfect myself in my mind. I thank my faith of love, which always teaches me to be patient, hopeful, joyful, and grateful. I thank all my friends, colleagues and the department faculty and staff for making my stay at College Station a memorable experience.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Tadhg P. Begley (advisor), Professor Frank M. Raushel and Professor Coran M. H. Watanabe of the Department of Chemistry and Professor Gregory D. Reinhart of the Department of Biochemistry and Biophysics.

The QDO mutant W153G, W153A, and S95G+H96G (Chapter 2) were provided by Dr. Hui Xu (the previous manager of the protein facility in Begley group). The DFT calculations related to the QDO reaction mechanisms (Chapter 3) were done by Dr. Lisa M. Perez (the manager of the Laboratory for Molecular Simulation at Texas A&M University). The compound, biotin-4-carboxybenzene sulfonyl azide (BiotinSA, **91**, Chapter 5) was synthesized by Dr. Sameh H. Abdelwahed (a senior post-doc in Begley group). All the plasmids related to the biosynthesis of LarA were kindly provided by Hausinger group at Michigan Station University (Chpater 5 and 6).

All other work included in this dissertation was completed by the student independently.

The funding sources are: Teaching Assistantship from the Department of Chemistry at Texas A&M University, Robert A. Welch Foundation under grant number A-0034 and National Institutes of Health (NIH) under grant number DK44083.

NOMENCLATURE

ABPP	Activity-based protein profiling
BiotinSA	Biotin-4-carboxybenzene sulfonyl azide
DanA	Dansyl azide
DFT	Density functional theory
Dha	Dehydroalanine
DMP	Discovery metabolite profiling
EIC	Extracted ion chromatography
IR	Infrared spectra
LarA	Lactate racemase
LC-MS	Liquid chromatography–mass spectrometry
Lp	Lactobacillus plantarum (or L. plantarum)
LRSA	Lissamine Rhodamine B Sulfonyl Azide
MSI	Mass spectrometry imaging
NaAD	Nicotinic acid adenine dinucleotide
NAD	Nicotinamide adenine dinucleotide
NiP2TMN	Nickel pyridinium-3,5-bisthiocarboxylic acid mononucleotide
NMR	Nuclear magnetic resonance spectra
P2CMN	Pyridinium-3,5-biscarboxylic acid mononucleotide
P2TMN	Pyridinium-3,5-bisthiocarboxylic acid mononucleotide
PCTMN	Pyridinium-3-carboxy-5-thiocarboxylic acid mononucleotide
PTM	Post-translational modifications

QDO	1-H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase
SAM	S-adenosyl-L-methionine
ThiS	Thiamine biosynthesis protein (A sulfur carrier protein)

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

This dissertation can be divided into two major parts. Firstly, I describe the trapping of the first case of enzyme-formed acylium ion active species with its native substrate. In addition, I also have developed small "click" probes to selectively detect the novel thiocarboxylated proteins and small metabolites generated in an NAD-derived cofactor biosynthesis pathway.

1.1 Trapping informative enzymatic reaction active intermediates

Enzymes catalyze chemical reactions in lives efficiently. It has been reported if there is no enzyme participation, biochemical reactions essential to life may take 2.3 billion years to complete[1]. Yet the mean value of the enzyme k_{cat}/K_M is about 10^7 M^{-1} s^{-1} and the k_{cat} is in the range of 10-1000 s^{-1}.[2] Our lab always has a strong interest in unveiling the mechanism of such astonishing biological catalyzing process. Considering enzymes as naturally evolved macromolecular catalysts, the biological reactions also follow the physical science rules. Based on these rules, we can propose multiple hypothetic reaction mechanisms which illustrate the stepwise transformation from the reaction substrate to its product reasonably but with the formation of different active intermediates and transition states in between. To figure out the mechanism of an enzymatic reaction, the most informative and convincing evidence is to trap the active intermediates experimentally.

Similar to reactions in solution, the common intermediates in enzymatic reactions also include cations, radicals, and anions. Cation cascades can commonly be found in the biosynthesis of terpene natural products[3]; radical SAM and P450 enzymes often catalyze the radical involved reactions[4]; interesting anion intermediates have been widely reported in reactions catalyzed by biotin and thiamin dependent enzymes.[5,6] A wide diversity of interdisciplinary methods was used

for directly and indirectly trapping of the enzymatic reaction intermediates. Some reaction intermediates have been directly "seen" by various spectroscopy techniques[7–9], and X-ray crystallography[10]. Synthetic chemistry has been used to make substrate analogs or isotope labeled substrates for informative in situ trapping.[11] More frequently, the active species were detected by measuring their characteristic stable derivatives. Many strategies were insightfully applied for this type of trappings, such as isotope labeling[12], radical clock chemistry[13], intermediate quenchers[14], protein engineering[15], and kinetic studies[16]. Computational methods also had broad applications on simulating the reasonable intermediate structures and evaluating their energy profiles.[17,18]

The first part of my Ph.D. dissertation was to trap the first reported case of enzyme-formed acylium ion intermediates (**6** and **7**) in a quinoline ring-cleavage reaction catalyzed by 1-H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase (QDO) (Figure 1.1). This acylium ion trapping challenged the literature proposed mechanism with a five-member cyclic peroxide (**5**) formation.[19] In Chpater 2 of this dissertation, I performed the experimental trapping of such cation intermediates. In Chapter 3, I described the computational modeling work (in collaboration with Dr. Lisa Perez from the Laboratory for Molecular Simulation at Texas A & M University) for a better understanding of this unique reaction mechanism in structure and energy aspects.

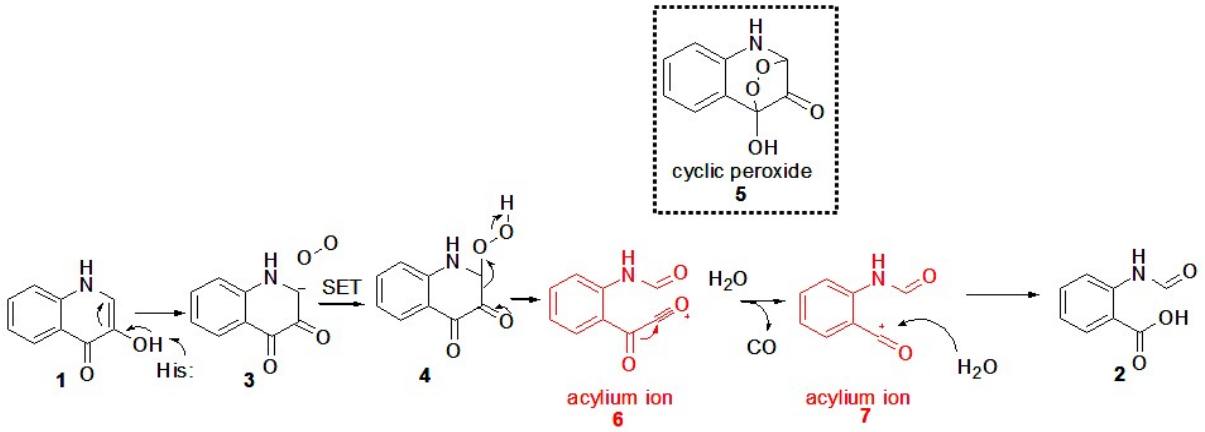


Figure 1.1: Our proposed mechanism of 1-H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase (QDO) catalyzed reaction. The intermediates **6** and **7** (labeled in red) are the experimentally trapped acylium ion intermediates. The literature reported cyclic peroxide intermediate (**5**) is in the dashed box.

1.2 Trapping thiocarboxylated proteins and small metabolites in the bacterial cell proteomes and metabolomes

Besides the reaction mechanism study, the enzyme function can be further investigated in the context of the whole cell. Initiated from simple molecules, the biosynthesis of the target natural products may require several steps of enzyme-catalyzed reactions. The relevant enzymes in a biosynthesis pathway of certain metabolite are often encoded in the same gene cluster (genes located next to each other within hundreds of base pairs) in bacterial cells. The characterization of an unknown enzyme may lead to the discovery of the whole biosynthesis pathway that it participates in. The rapid development in bioinformatics has dramatically accelerated this process. Some enzymes sharing similar gene sequences (homologs) may catalyze different types of reaction, such as the two radical SAM enzymes, ThiC and BzaF.[20] On the other hand, ThiS-like proteins are found to be sulfur carriers in different biosynthesis pathways.[21] Since the bioinformatic "in silico" annotation of hypothetical proteins (based on their gene sequence

similarity with the known proteins) may cause some errors in determining their real functions, methods that can directly "trap" the target enzymes or their related metabolites in the cell extract with high specificity and sensitivity are desired to complement the drawbacks of the bioinformatic analysis.

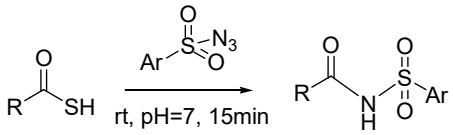
Post-translational modifications (PTMs) increase the proteome diversity. Some enzymes must be activated by certain kind of PTM. If the PTM moiety of the enzyme can be specifically trapped by delicately designed probes in the proteome, an activity-based protein profiling (ABPP)[22] can be set up for many applications such as enzyme activity dynamic monitoring, enzyme regulation/inhibition studies, and diseases diagnosis. Some cofactors must covalently bind to the enzyme for its activity reconstitution, which can also be considered as an activity-dependent PTM. The selective PTM labeling may also contribute to the discovery of novel PTM sites in the proteome which may indicate the gain of new protein functions.

Metabolomics is a relatively new concept introduced to "omics" sciences. The bacterial cell metabolome refers to the complete set of low molecular weight compounds in the bacterial cell extract, which contains the substrates and products of enzymatic reactions. Compared with other omics research, metabolomics has more impact on the phenotype rather than the genotype of the cells. High-resolution LC-MS is a dominant tool for metabolome research. The Cravatt's group has developed the discovery metabolite profiling (DMP) method to identify the endogenous substrates of enzymes.[23] The mass spectrometry imaging (MSI) techniques have also been broadly used in metabolomic studies to illustrate the interactions between living species and environment, and the disease diagnosis.[24] The Siuzdak's group has developed the LC-MS information based database, XCMS Online metabolomics platform, for system biology studies.[25] Some metabolites of interest may degrade readily or have poor

behavior for LC-MS detection. By selective chemical derivatization, such compounds are possible to be trapped successfully.

Protein thiocarboxylation is a PTM that was previously only found in ThiS-like sulfur carrier protein superfamily.[26] The Hausinger's group reported a unique NAD-derived thiocarboxylate cofactor that covalently bound to the lactate racemase (LarA) in *Lactobacillus plantarum*.[27] This cofactor can be regarded as a new type of thiocarboxylation PTM of LarA. During its biosynthesis pathway (LarB, LarE and LarC dependent), several protein thiocarboxylates and small thiocarboxylate metabolites were formed (Figure 1.2), among which I reported the selectively "click" chemical trapping of two protein thiocarboxylates, LarA-COSH (**18**) and LarE-COSH (**14**) in the bacterial proteomes (Chapter 5) and two small thiocarboxylate metabolites, **12** and **15**, in the bacterial metabolomes (Chapter 6).

A)



R = proteins (**14** and **18**) or small molecules (**12** and **15**)
Ar can link to a fluorescent or a biotinated group

B)

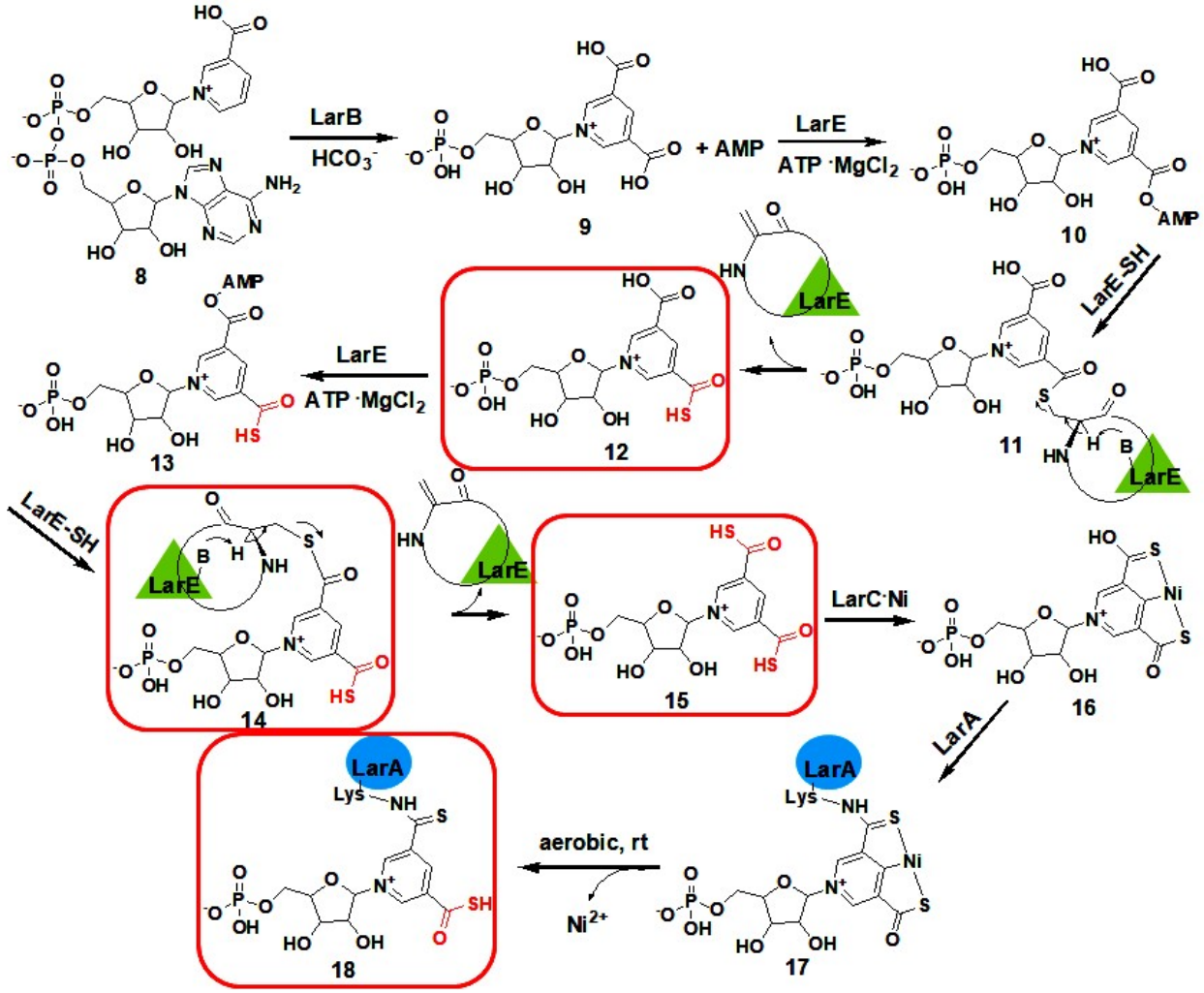
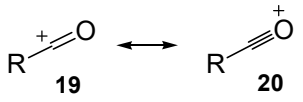


Figure 1.2: A) The key "click" reaction for the thiocarboxylates specific trapping; B) The LarA cofactor biosynthesis pathway. The protein thiocarboxylates (**14** and **18**) and small thiocarboxylate metabolites (**12** and **15**) were selectively labeled in the bacterial cell extract (in red boxes). The Cys176 of LarE donated the sulfur to the cofactor and formed dehydroalanine (Dha).

2. ENZYME-CATALYZED ACYLIUM ION FORMATION

2.1 Acylium ions

Acylium ions are resonance-stabilized cations (Figure 2.1). They are important active intermediates in organic reactions, such as Friedel-Crafts acylation[28], acid-catalyzed esterification, and hydrolysis.[29] In 1943, Seel isolated and characterized the first stable acylium ion salt from a reaction of acetyl fluoride and BF_3 at a low temperature.[30] The acylium ion formation could be determined by nucleophile capture (commonly used for carbocation trapping), kinetic measurements, electronic absorption spectra (UV-vis spectra), infrared spectra (IR), nuclear magnetic resonance spectra (NMR), electron spectroscopy, X-ray crystallography and mass spectrometry.[31–33] Density functional theory (DFT) computational modeling was also used to rationalize the formation of an acylium ion by calculating the reaction energy profile.[34] However, to our knowledge, the acylium ion has never been previously reported to be formed in a native enzyme-catalyzed reaction. The experimental data in this chapter provided strong evidence in trapping the first case of the acylium ion generated from a metal/cofactor independent dioxygenase enzyme, 1-H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase (QDO), catalyzed reaction.



19 **20**

Figure 2.1: The resonance structures of aromatic acylium ions.

2.2 Metal/cofactor independent oxygenase

Molecular oxygen (O_2) is a strong oxidant which abundantly exists in today's atmosphere (~21%). It usually maintains the stable triplet form. Aerobic organisms have evolved several metal complexes and organic cofactors (Figure 2.2) to activate the triplet O_2. Oxygenases can incorporate one or both oxygen atom(s) into their substrates. Most oxygenases are iron, copper, flavin or pterin cofactor dependent enzymes.[35]

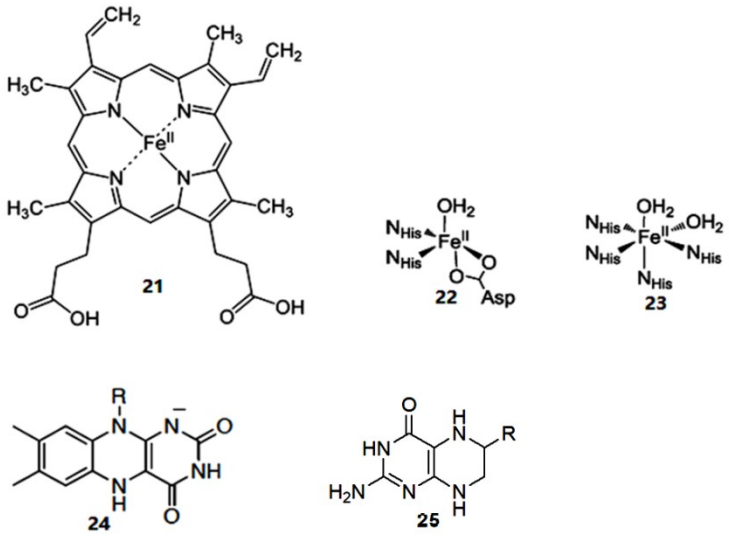
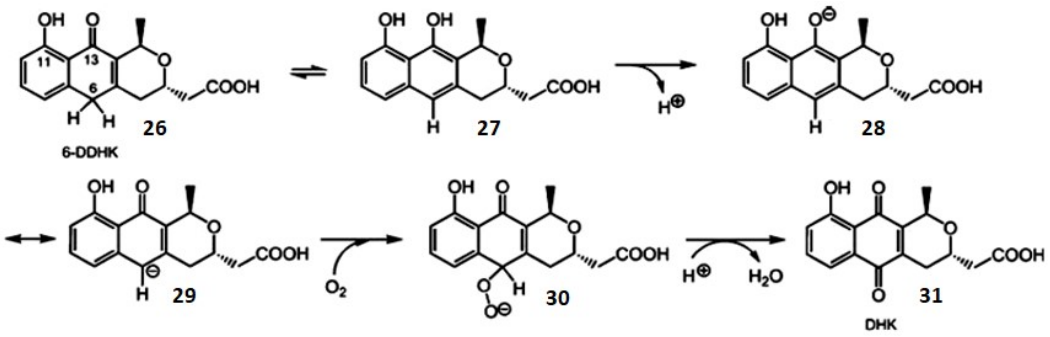


Figure 2.2: Examples of the metal and organic cofactors involved in O_2 activation: heme (**21**), non-heme iron (II) centers (**22** and **23**), reduced flavin (**24**), and pterin (**25**)

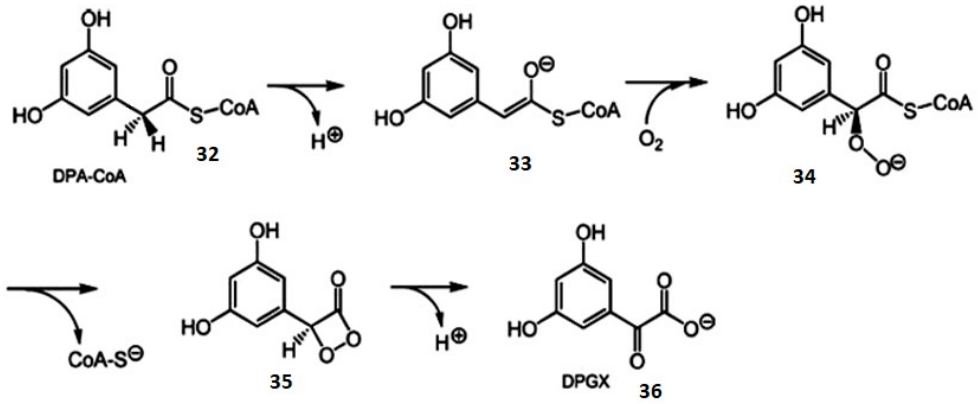
Since the 1990s, several enzymes from different protein superfamilies have been reported to catalyze the oxygenation reaction without the requirements of either metal or organic cofactors. Besides QDO studied in this chapter, other examples included ActVA-Orf6, DpgC, and RLuc[36] (Figure 2.3). ActVA-Orf6 is a monooxygenase which oxidizes 6-deoxydihydrokalafungin (6-DDHK) to dihydrokalafungin (DHK) in the biosynthesis of the polyketide antibiotic actinorhodin from *Streptomyces coelicolor*. DpgC with the full name, 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase, participates in the biosynthesis of vancomycin and

teicoplanin antibiotics from *Amycolatopsis orientalis*. RLuc is a cofactor independent lusiferase from *Renilla reniformis* (sea pansy).

A) ActVA-Orf6



B) DpgC



C) RLuc

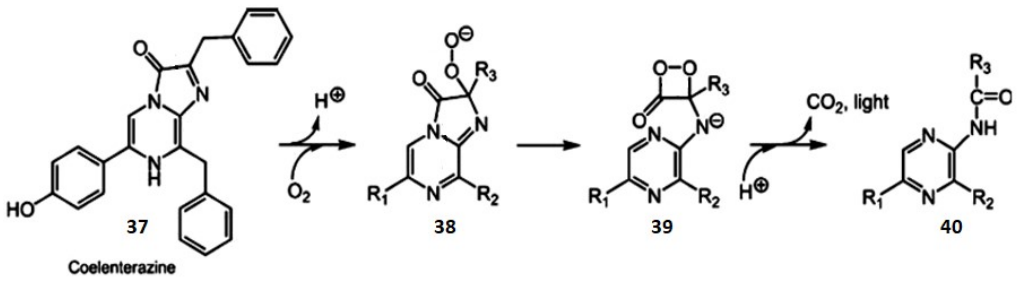


Figure 2.3: Proposed O_2 incorporation mechanisms of three reported metal/cofactor independent oxygenases, ActVA-Orf6, DpgC, and RLuc.

These cofactor-independent oxygenases manage to mediate the direct single-electron transfer from a substrate anion to dioxygen. This process has some similarity to the O_2 activation by the reduced flavin (Figure 2.4).

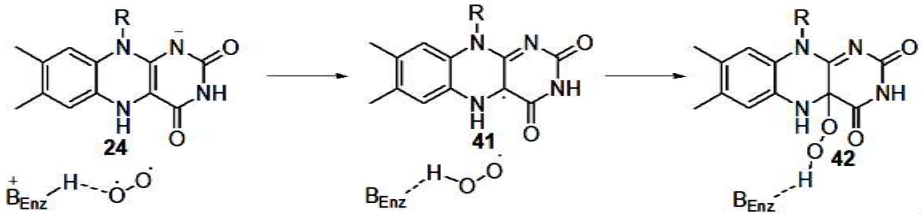


Figure 2.4: Proposed O_2 activation mechanism by the reduced flavin.

2.3 1-H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase (QDO) in the quinoline degradation

Quinoline derived compounds are carcinogens and mutagens that are widespread in the coal tar contaminated water and soil, as well as the waste generated from the pharmaceutical manufacturing. To develop an efficient and environment-friendly way to reduce the toxicity of such harmful waste, bacteria strains that can biodegrade the quinoline compounds have been isolated. In 1990, a 1H-4-oxoquinoline (**43**) catabolic pathway has been found from *Pseudomonas putida 33/1*(Figure 2.5).[37,38] 1-H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase (QDO) is a cofactor-independent dioxygenase catalyzing a remarkable N-heteroaromatic ring-cleavage reaction. It uses 1-H-3-hydroxy-4-oxoquinoline (**1**) and O_2 as substrates and produces N-formylanthranilate (**2**) and CO (Figure 2.6A). This is a key reaction in 1H-4-oxoquinoline degradation.

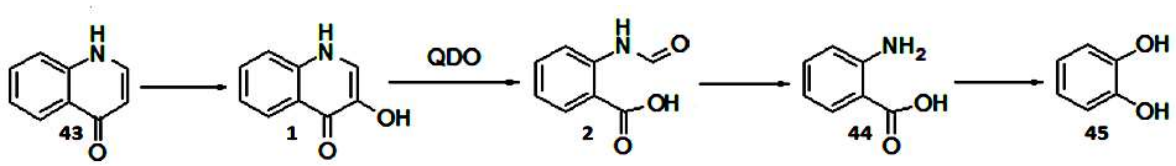


Figure 2.5: The 1H-4-oxoquinoline degradation pathway in *Pseudomonas putida 33/1*.

QDO belongs to the α/β-hydrolase fold superfamily which has the conservative structure of Serine 95/Histidine 244/Aspartic acid 120 catalytic triad[19] (Figure 2.6B, in green stick). Yet the Serine 95 residue does not function as a nucleophile as reported in most hydrolases. Instead, it only facilitates the substrate binding through a hydrogen-bonding interaction. The S95A mutant still maintains 90% QDO activity, while H244A or D126A almost demolishes the QDO activity. Therefore, H244 and D120 are sufficient to form a general base pair to catalyze the initial substrate deprotonation step. QDO has a homolog, 1H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase (HOD), from *Arthrobacter sp. Rü61a*. It catalyzes the same ring cleavage reaction as QDO does but using 1H-3-hydroxy-4-oxoquinaldine (**46**) as the substrate. HOD shares only 36% amino acid identity with QDO, but their secondary protein structures overlap very well (Figure 2.6C). HOD+substrate complex structure shows the H251, D126 and substrate position clearly (Figure 2.6B, in cyan stick). The QDO+substrate (**1**) complex structure is not available. In Chapter 3, computational docking of the substrate into the QDO active site is shown. The H244/D120 (H251/D126 in HOD) general base pair can deprotonate 3-OH and form a resonance anion at the C2 position for further O_2 incorporation (Figure 2.6D).

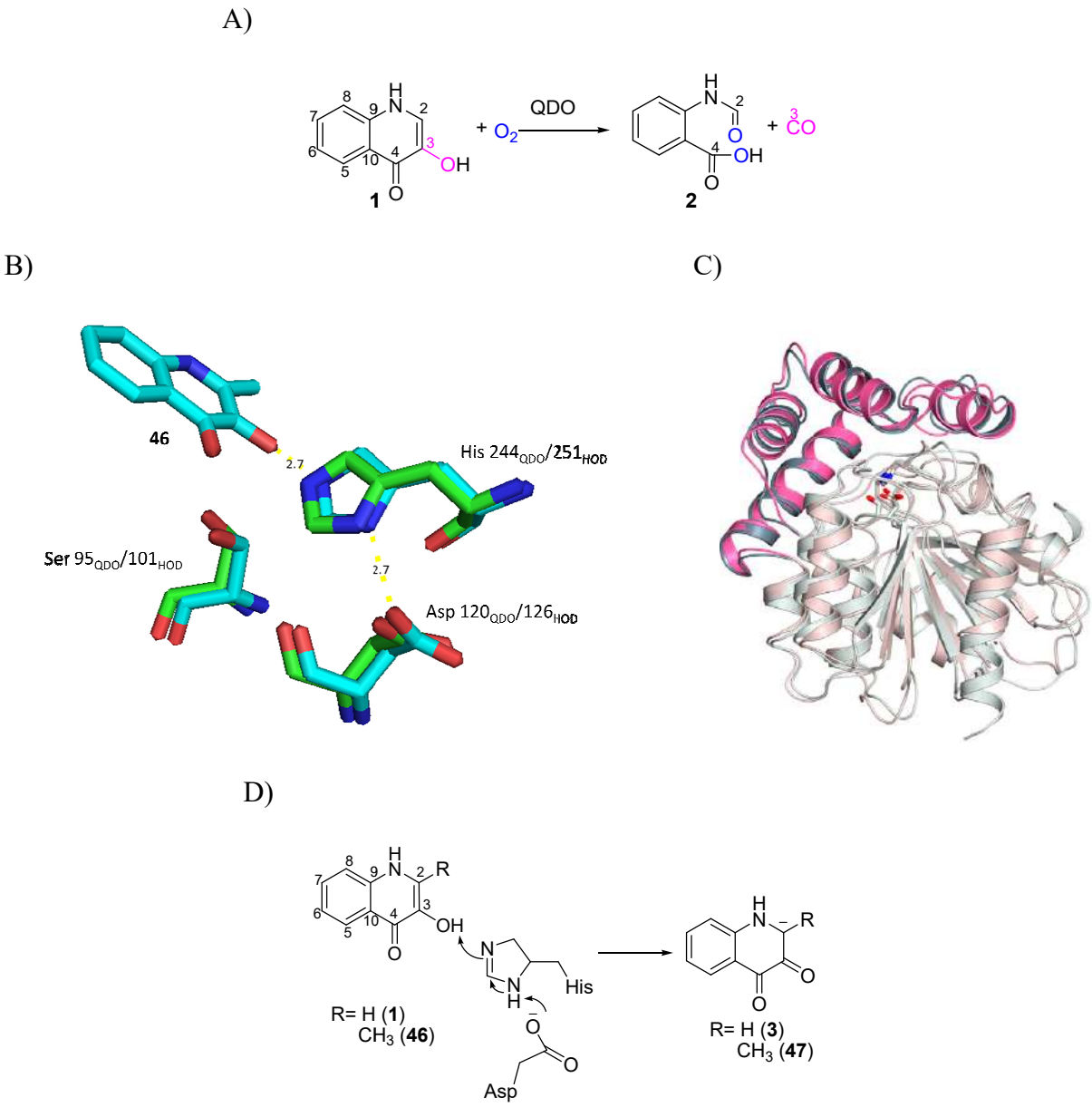


Figure 2.6: QDO/HOD catalyzed reactions and their active site structures. A) The QDO catalyzed reaction. B) The Ser/His/Asp catalytic triads overlay of QDO (in green stick, PDB: 3IBT) and HOD (in cyan stick, PDB: 2WJ4) structures. C) The cartoon overlay of QDO (in pink, PDB: 3IBT) and HOD (in gray, PDB: 2WJ3) crystal structures. D) The initial Asp/His catalyzed step of the substrate (**1** or **46**) deprotonation.

2.4 Metal-dependent systems catalyze the reactions similar to QDO

QDO catalyzes a remarkable N-heteroaromatic ring cleavage reaction. Its gaseous product, carbon monoxide (CO), is an important signaling and poisoning molecule that commonly exists in living organisms.[39–42] Besides the substrate-assistant metal/cofactor independent fashion, nature has also developed various versions of metal-dependent dioxygenases to catalyze the same type of reaction as QDO does. A typical example is the quercetin 2,3-dioxygenase.

Quercetin is a plant originated flavonol with antimicrobial and antioxidant activity. It has a similar structure to 1-H-3-hydroxy-4-oxoquinoline, and can be degraded by certain bacterial and fungal strains in a way similarly to the QDO chemistry. The corresponding quercetin 2,3-dioxygenase is metal-dependent and has no structural resemblance to QDO. The first quercetin 2,3-dioxygenase was discovered in *Aspergillus* fungal strains in 1971. It has a copper center which can either directly activate the O_2 first to react with the quercetin substrate, or activate the quercetin first followed by the O_2 addition (Figure 2.7). Quercetin 2,3-dioxygenases in bacteria can have more metal tolerance. It has been reported that they could use a diversity of divalent metal ions such as Fe(II), Co(II), Ni(II) and Mn(II).[43] Metal complex biomimetics of quercetin 2,3-dioxygenase have also been widely reported.[44–46]

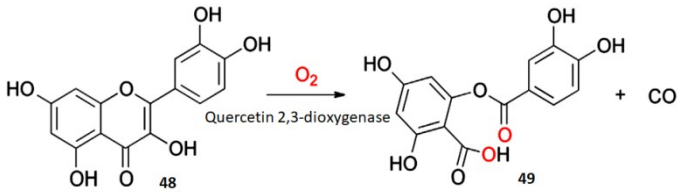
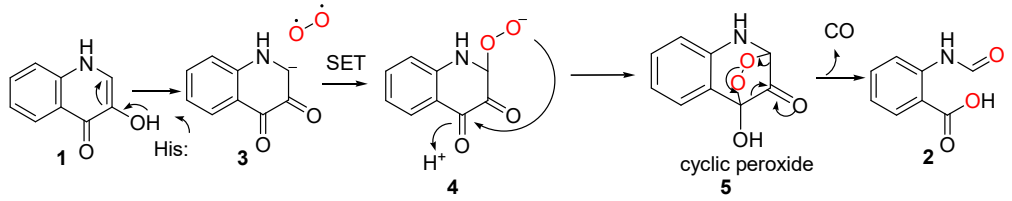


Figure 2.7: The quercetin 2,3-dioxygenase catalyzed reaction

A)



B)

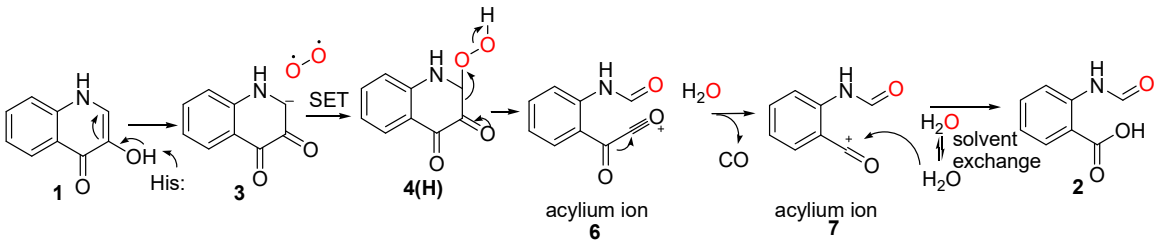


Figure 2.8: Alternative mechanisms of QDO reaction. A) The reported mechanism with the cyclic peroxide (**5**) formation; B) our proposed mechanism with the acylium ions (**6** and **7**) formation. $^{18}O_2$ is marked in red in $^{18}O_2$ isotope labeling reaction.

2.5 QDO catalyzed acylium ion formation

Fetzner group has proposed a five-member ring cyclic peroxide (**5**) as the key intermediate of the QDO reaction[19] (Figure 2.8A). Yet, **5** has not been experimentally trapped over a decade. In this chapter, we proposed an alternative mechanism (Figure 2.8B) involving the sequential formation of two acylium ions (**6** and **7**) active species during the QDO enzymatic catalysis process. When $^{18}O_2$ (Figure 2.8, in red) is incorporated to **1** through an aspartic acid-histidine general base catalytic process to form the peroxide **4(H)**, followed by the $^{18}OH^-$ releasing and the formation of acylium ion **6**. The CO then readily releases from **6** to produce the more stable acylium ion **7**. The $^{18}OH^-$ then rebinds to **7** to generate the product **2**. If the $^{18}OH^-$ could undergo solvent-exchange with the ^{16}O water fast enough at the QDO active site, the mono-^{18}O labeled product **2** may be observed. On the other hand, the ring cleavage of **5** could only result in the di-^{18}O labeled **2** formation. Therefore, trapping the mono-^{18}O labeled product **2**

could provide unambiguous evidence indicating the acylium ion **7** formation. It has been reported that the wild-type QDO full reaction with its natural substrate **1** under $^{18}O_2$ could only produce **2** with di-^{18}O labeled but no mono-^{18}O labeled at all.[47] If our hypothesized acylium ion mechanism is correct, the absence of the mono-^{18}O labeled **2** could be due to the much slower solvent exchange rate at the active site than the hydroxide rebinding rate. Experiments were designed to either stabilize the acylium ions in order to slow down the hydroxide rebinding rate or to increase the solvent exchange rate at the QDO active site. The acylium ion **6** might also be trapped by the solvent water before its CO releasing and form the keto acid side product (Figure 2.9A).

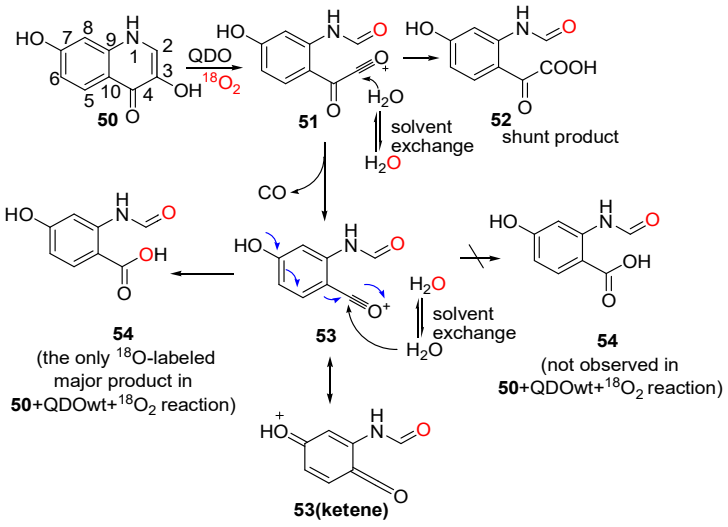
2.6 Result and discussion

2.6.1 Trapping the first generated acylium ion (51, an analog of 6, see Figure 2.8 and 2.9) using a substrate analog, 3,7-dihydroxyquinolin-4(1H)-one (50, see Figure 2.9)

Originally to stabilize the acylium cation **7** for efficient solvent trapping, the substrate analog, 3,7-dihydroxyquinolin-4(1H)-one (**50**, Figure 2.9) with an electron-donating hydroxyl group at C7 of its phenyl ring was synthesized. It was a substrate of QDO and produced **54** (analog of **2**, see Figure 2.8 and 2.9) as the major product. However, it did not help observing the solvent capture of the acylium ion **53** (analog of **7**, see Figure 2.8 and 2.9). Under $^{18}O_2$, we could only observe the formation of the di-^{18}O labeled **54**. No mono-^{18}O labeled **54** was detected. Nevertheless, the high-resolution LCMS analysis showed that in addition to the expected major product **54** eluted at 9.5min, a new minor product peak was found at 8 min with an [M-H] mass of 208.0 Da (Figure 2.9B, the red arrowed peak). This product was demonstrated to form enzymatically by a time-course reaction with fixed enzyme concentration and by varying enzyme

concentration in the same reaction time period (Figure 2.10). Since its mass was 28 Da higher than the major product **54** ([M-H] = 180), we proposed the structure of this new minor product as **52** (Figure 2.9A).

A)



B)

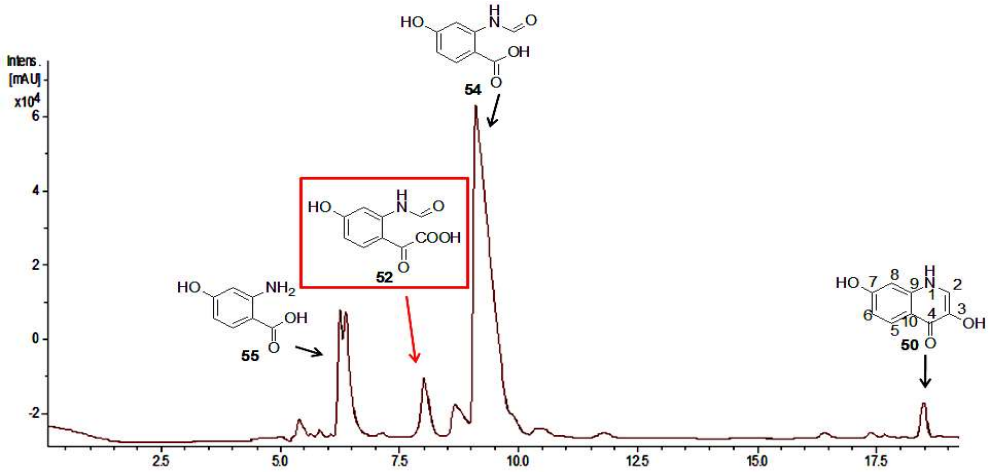


Figure 2.9: The acylium ion **51** (analog of **6** in Figure 2.8B) was trapped by solvent water to form the shunt product **52**. A) The proposed mechanism of the solvent trapping of **51** under $^{18}O_2$ (in red). The acylium ion **53** (analog of **7** in Figure 2.8B) solvent trapping product (the mono-^{18}O labeled **54**) was not observed. B) LC analysis of **50** + QDOwt full reaction mixture at 254 nm. The new shunt product **52** peak was in the red box. **55** was the formyl-hydrolyzed product of **54**.

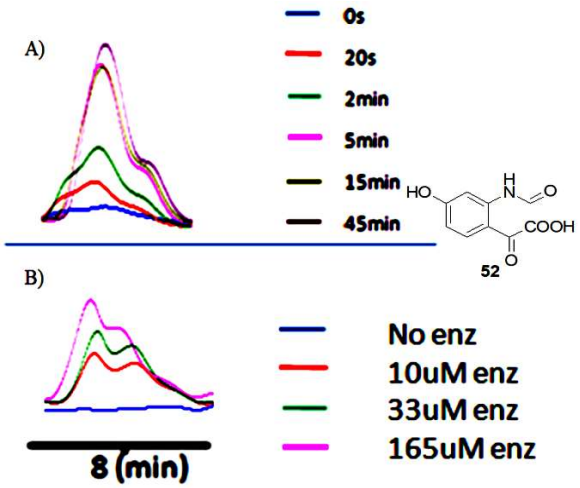


Figure 2.10: 52 was formed enzymatically determined by A) the time-course of reaction with a fixed concentration of QDOwt, 100 μM; and B) variation of QDOwt concentrations in the same reaction time period, 15 minutes. The starting concentration of **50** was the same, 3 mM, in both A) and B) tests. The traces were the HPLC analysis at 254 nm. The peak doublets were due to the rotamers formation of **52**. **50** did not form **52** non-enzymatically at pH 6.5-7 buffer solution for 2 days.

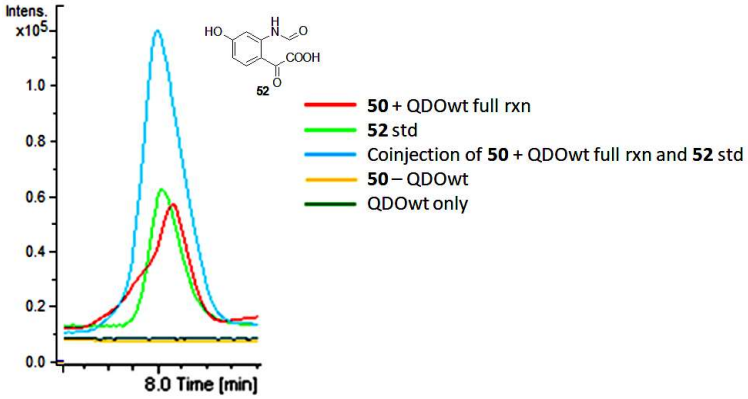


Figure 2.11: LC-MS coelution of the shunt product **52** peak at 8 min with the synthesized authentic standard. The traces were EIC = 208.0 Da (the [M-1] of **52**) in the negative mode.

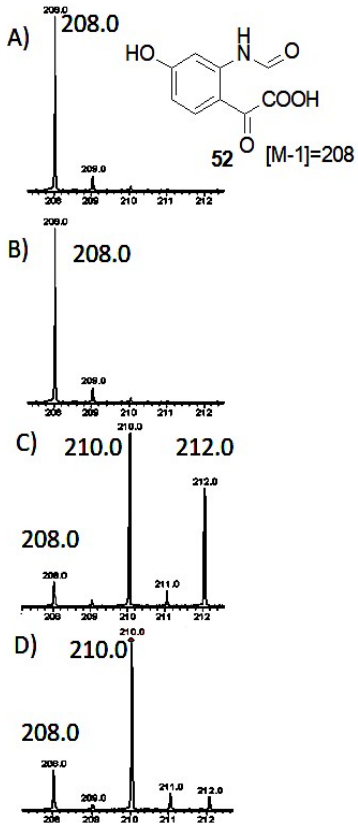


Figure 2.12: MS analysis of **52** generated from QDOwt+**50** full reaction. A) MS of **52** synthetic standard ([M-1] = 208.0). B) MS of **52** formed from QDOwt+50 full reaction in non-label conditions. C) MS of **52** formed from QDOwt+**50** full reaction under $^{18}O_2$. D) MS of **52** formed from QDOwt+**50** full reaction in $H_2{}^{18}O$.

The **52** standard was then synthesized, and it coeluted well with the new minor product peak through the LCMS analysis (Figure 2.11). More interestingly, the reaction was performed in $^{18}O_2$ and $H_2{}^{18}O$ and clearly showed the solvent-exchange mono-^{18}O labeled **52** formation with the [M-H] mass 210.0 Da (Figure 2.12C-D). The ^{18}O labeled positions were figured out through the MS/MS fragmentation analysis (Figure 2.13 and Figure 2.14) which matched the proposed labeling patterns very well (Figure 2.9A). When the reaction was under $^{18}O_2$, the mono-^{18}O labeled minor product **52** had ^{18}O labeled exactly at formyl aldehyde position as proposed. The

reaction in $H_2{}^{18}O$ produced the mono-^{18}O product **52** only with ^{18}O labeled at the carboxylic acid position.

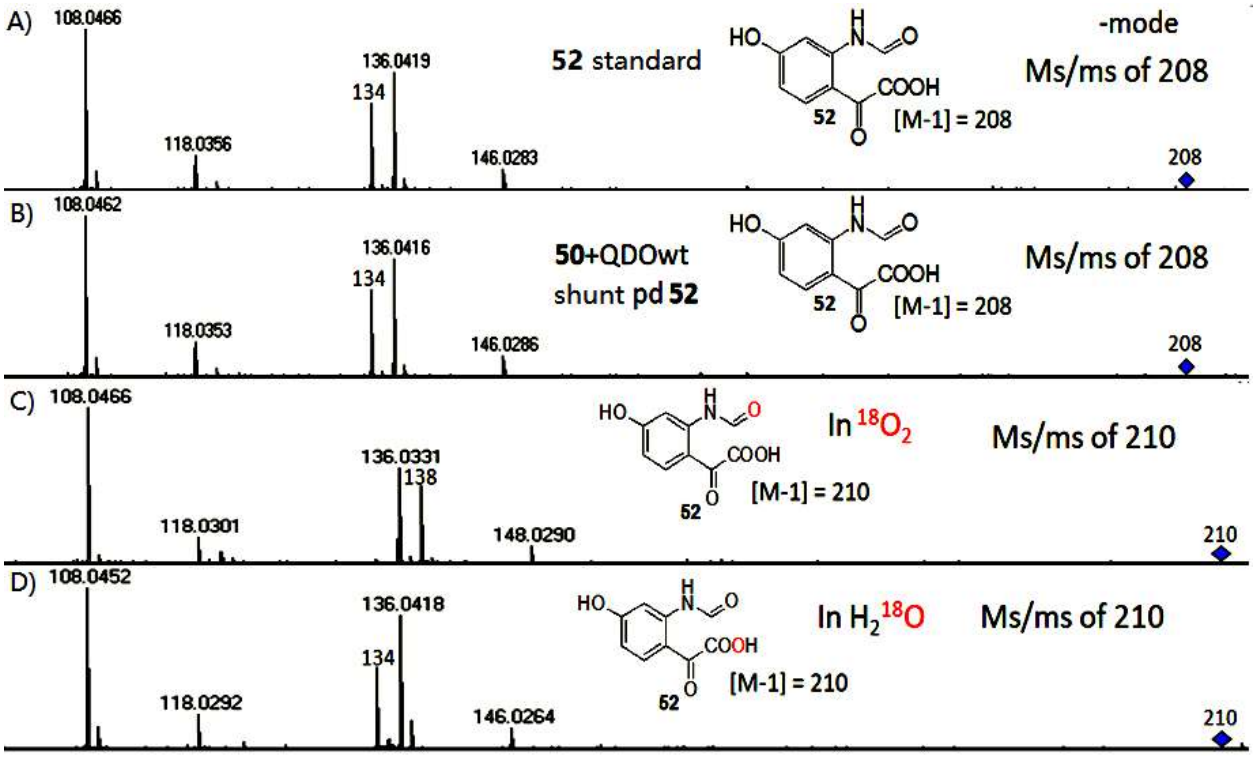
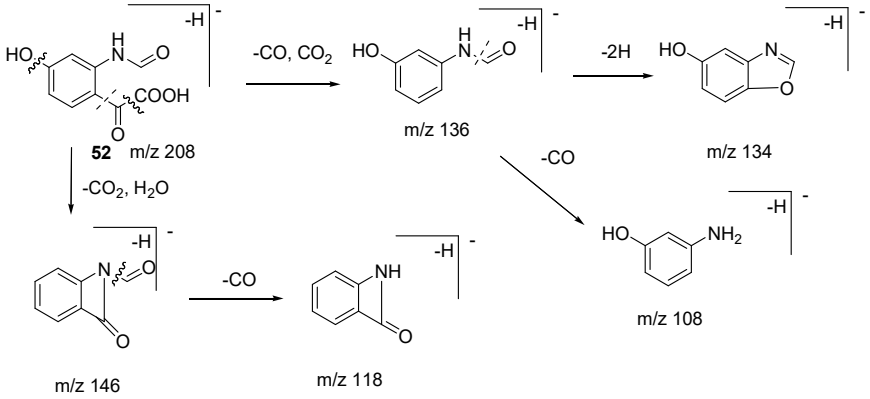
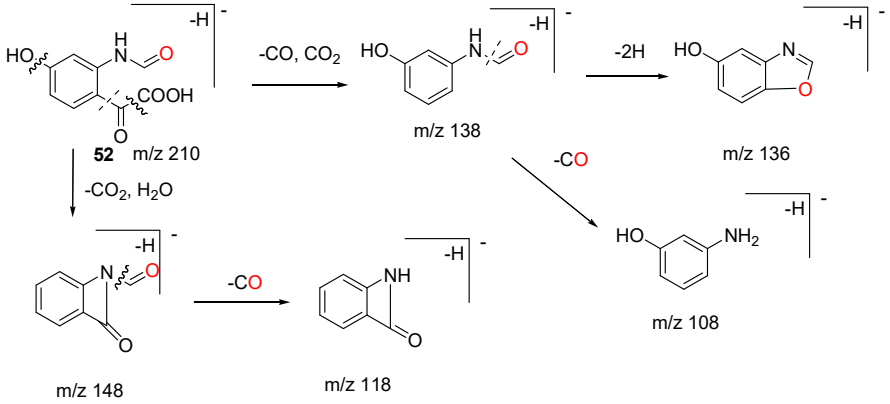


Figure 2.13: MS/MS data (in negative mode) of **52** generated from A) the synthetic standard (the MS/MS patterns maintained the same when dissolving the **52** standard in both $H_2{}^{16}O$ and $H_2{}^{18}O$ for 2 days); B) **50**+QDOwt full reaction in non-label conditions; C) **50**+QDOwt full reaction in $^{18}O_2$ (^{18}O is in red); D) **50**+QDOwt full reaction in $H_2{}^{18}O$. (See analysis in Figure 2.14)

A) The MS/MS of **52** standard in both $H_2{}^{16}O$ and $H_2{}^{18}O$ for 2 days and the product **52** from non-labeled **50**+QDOwt full reaction



B) The MS/MS of product **52** from **50**+QDOwt full reaction in $^{18}O_2$



C) The MS/MS of product **52** from **50**+QDOwt full reaction in $H_2{}^{18}O$

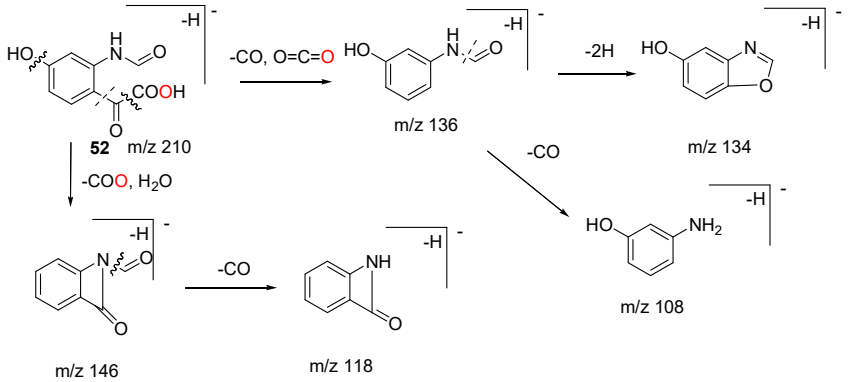


Figure 2.14: MS/MS analysis summary of **52**. A) The MS/MS fragmentation of **52** standard in both $H_2{}^{16}O$ and $H_2{}^{18}O$ for 2 days and the product **52** from non-labeled **50**+QDOwt full reaction. B) The MS/MS fragmentatation of product **52** from **50**+QDOwt full reaction in $^{18}O_2$. C) The MS/MS fragmentation of product **52** from **50**+QDOwt full reaction in $H_2{}^{18}O$

Controls had been done to prove that neither **50** nor **52** had any non-enzymatic solvent-exchange occurring even after 2 days incubation with $H_2{}^{18}O$ (Figure 2.15C-D). All these results provided the evidence of successful trapping the first-generated acylium ion intermediate **51** (analog of **6**) rather than the CO-released acylium ion **53** (analog of **7**) in the mechanism shown in Figure 2.9A.

The substrate analog **50** was designed to stabilize the acylium ion **53** by forming a ketene resonance form, **53(ketene)** (Figure 2.9A). However, the C7 hydroxyl electron-donating effect of **50** seemed actually stabilizing the C-C bond between C3 and C4, which slowed down the CO releasing rate and increased the chance of the water attacking to the acylium ion **51** to form the keto carboxylic acid **52** as a shunt product.

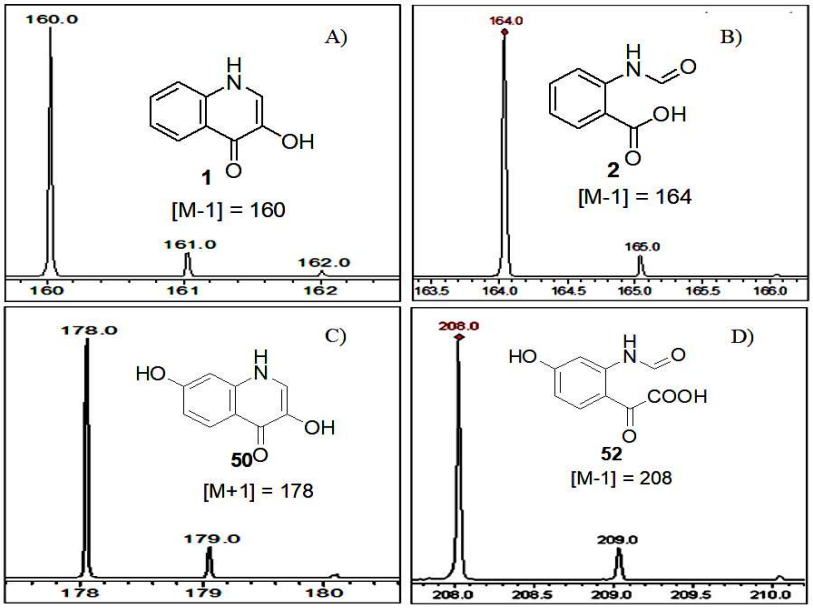


Figure 2.15: 1, 2, 50, 52 had no non-enzymatic solvent-exchange in $H_2{}^{18}O$ for 2 days. A) MS of **1** std in $H_2{}^{18}O$ for 2 days with [M-H] = 160.0 Da. B) MS of **2** std in $H_2{}^{18}O$ for 2 days with [M-H] = 164.0 Da. C) MS of **50** std in $H_2{}^{18}O$ for 2 days with [M+H] = 178.0 Da. D) MS of **52** std in $H_2{}^{18}O$ for 2 days with [M-H] = 208.0 Da.

We further investigated this controversy between our proposal and the experimental data by computational calculations (Figure 2.16A-B, details were described in Chapter 3). In solution, the calculated CO releasing energy barrier of the 7-OH substituted substrate derived acylium ion **51** was only ~1 kcal/mol lower than the native substrate derived acylium ion **6**, which was within the $\pm$ 5 kcal/mol error range. This calculation suggested that the 7-OH substitution did not affect the CO releasing rate significantly. The 7-OH electron-donating effect made the C4 position of **53** less eletrophilic to quickly attract the hydroxide generated from O_2 inside the binding pocket, so the hydroxide might have enough time for solvent-exchange before binding to **53**. However, there was also a competition between C3 and C4 of **51** for the hydroxide attacking since they were both electron-deficient carbonyl carbons. When the C4 became less electron-deficient, it also increased the chance for the hydroxide to attack C3 and formed the shunt product **52**. In addition, 7-OH substitution of **50** was bulkier than 7-H in **1**, which might affect the binding of **50** to let the hydroxide more available to C3. So the rate of the hydroxide binding to C3 might be competitive against the CO releasing rate and the accumulation of **52** was observed. Besides, the C3 of **51** might be "squeezed" closer to the entrance of the binding pocket to gain more solvent exposure (Figure 2.16C) than **6**. There was a competition between the O_2-originated hydroxide (inside the binding pocket) and solvent water (outside the binding pocket) to attack the C3 of **51**. Thus both mono-^{18}O labeled and di-^{18}O labeled **52** were observed. After CO releasing, the acylium ion **53** might be small enough to bury inside the binding pocket, which prevented a fast solvent-exchange occurring before the O_2-oringated hydroxide binding to form **54**. It was not surprised to see that only the di-^{18}O labeled **54** generated.

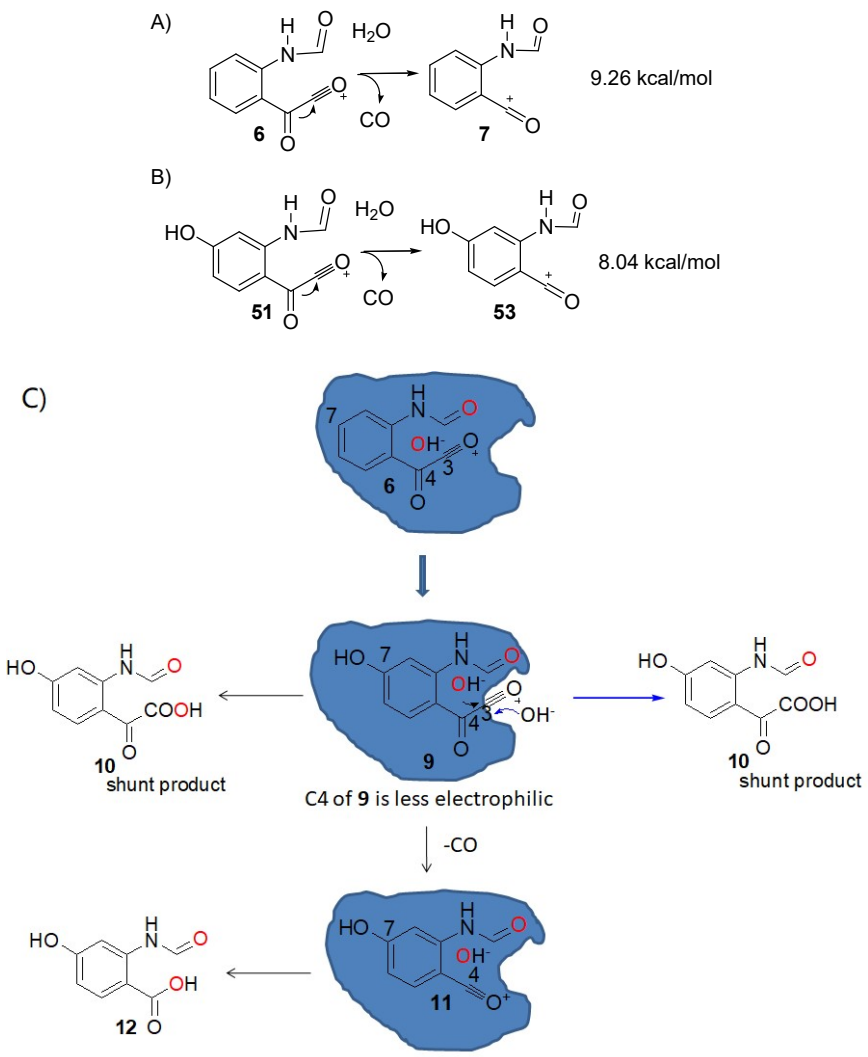


Figure 2.16: The CO releasing step energy barrier comparison between A) the native substrate derived acylium ion **6** (9.26 kcal/mol) and B) the 7-OH substituted substrate derived acylium ion **51** (8.04 kcal/mol). C) The cartoon of the QDO protein binding pocket (in blue) that led to the formation of only solvent-exchanged mono ^{18}O labeled **52** but no mono ^{18}O labeled **54** was observed.

2.6.2 Trapping acylium ion 7 (Figure 2.8B)

Crystal structure of the QDO+**1** complex is not available. Its homolog, 1-H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase (HOD), catalyzes the same type of reaction with the substrate,

1-H-3-hydroxy-4-oxoquinaldine (**46,** Figure 2.6D and Figure 2.17B). HOD+**46** complex crystal structure (PDB: 2WJ4) suggests that this enzyme have a very tight binding pocket which may prevent a fast solvent-exchange at the active site (Figure 2.17A).

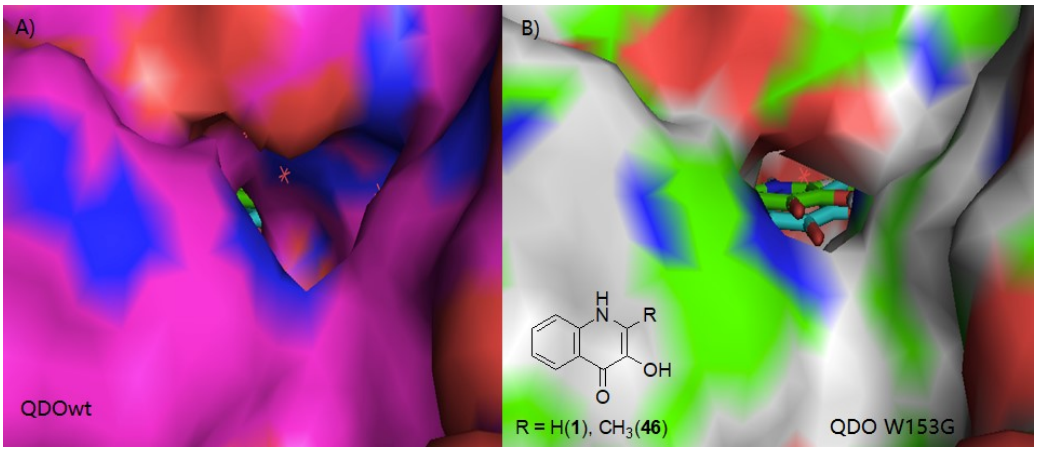


Figure 2.17: A) The surface view of the entrance of QDOwt binding pocket (PDB: 3IBT); B) The same viewing angle of the optimized QDO W153G binding pocket entrance. The docked QDO substrate **1** inside the binding pocket was in green stick, and the overlapped HOD substrate **46** was in cyan stick.

Site-direct mutants	mono-^{18}O: di-^{18}O labeled pd **2**
W153G	~1:6
W153A	~1:40
S95G&H96G	<1:400
S95G	<1:400
H244G	<1:400
Q32A	<1:400
H96G	<1:400

2
[M-1] = 164

Table 2.1: Table of designed site-directed mutants and their mass peak intensity ratios of mono-^{18}O labeled product **2** ([M-1]=166) : di-^{18}O product **2** ([M-1]=168) under $^{18}O_2$. Only QDO W153G and W153A mutants catalyzed reaction with **1** formed the mono-^{18}O labeled product **2**

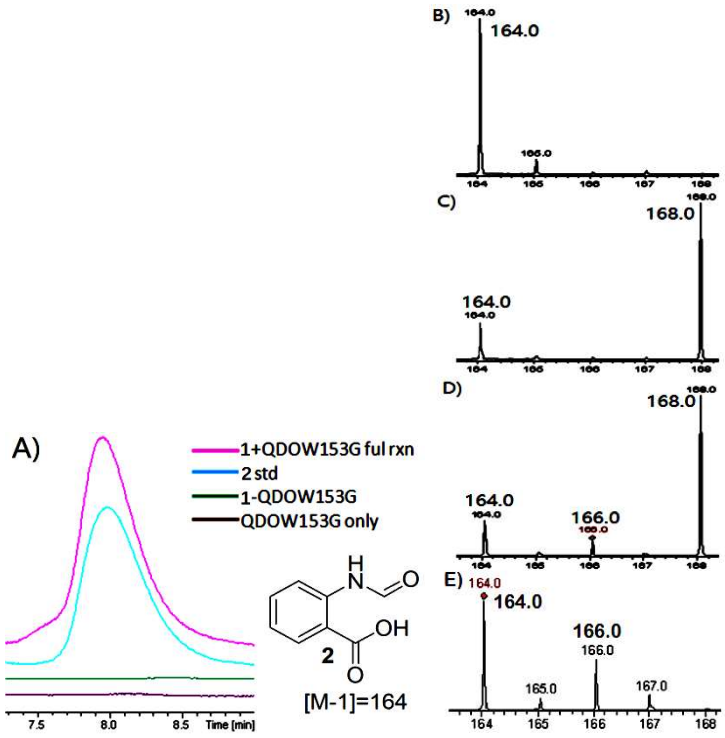


Figure 2.18: The solvent trap of acylium ion **7** determined by LCMS. A) The LC traces of the product **2** from **1**+QDOW153G full reaction at 254 nm. B) The mass of product **2** from the non-labeled **1**+QDOW153G full reaction. C) The mass of product **2** from the **1**+QDOwt full reaction in $^{18}O_2$. D) The mass of product **2** from the **1**+QDOW153G full reaction in $^{18}O_2$. E) The mass of product **2** from the **1**+QDOW153G full reaction in $H_2{}^{18}O$.

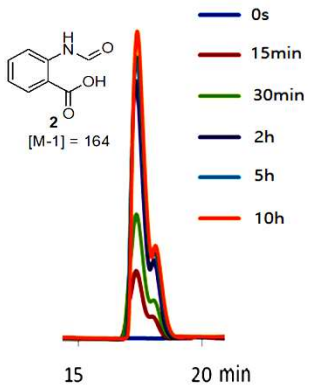


Figure 2.19: 2 was formed enzymatically in QDO W153G+**1** full reaction (pH 6.5-7) in 10 h time course. The peak doublet was due to the rotamers of **2**. **1** did not form **2** non-enzymatically at pH 6.5-7 in 2 days.

To increase the solvent-exchange rate at QDO active site, attempts to open up the binding pocket entrance had been made by designing a series of site-direct mutants (Table 2.1). Under $^{18}O_2$, the solvent-exchanged mono-^{18}O product **2** with the $[M-H]^-$ mass 166.0 Da was finally observed when mutating Trp153 to the small side-chain residues, Ala and Gly (Table 2.1 and Figure 2.18). The mono-^{18}O product **2** was generated enzymatically at pH = 6.5~7 (Figure 2.19).

The incubation of **1** and **2** in $H_2{}^{18}O$ separately at room temperature for 2 days did not produce any solvent-exchange **1** or **2** (Figure 2.15A-B). The LC-MS/MS fragmentation confirmed that the ^{18}O was labeled at formyl aldehyde position (Figure 2.20D and Figure 2.21D). The QDO W153G+**1** reaction in $H_2{}^{18}O$ also showed solvent-exchanged mono-^{18}O product **2** formation with ^{18}O labeled at carboxylic acid position (Figure 2.20E and Figure 2.21C), which was consistent with the reaction in $^{18}O_2$. These results provided strong evidence in successful solvent-trapping of the second acylium ion **7** in our proposed mechanism. Quantitatively, W153G produced more solvent-exchanged mono-^{18}O product **2** than W153A (Table 2.1). In another word, a smaller side-chain residue in replacement of W153 could make the QDO binding pocket open up more for better solvent-exchange efficiency. The MM optimized structure of QDO W153G showed a significant open-up of its binding pocket (Figure 2.17B)

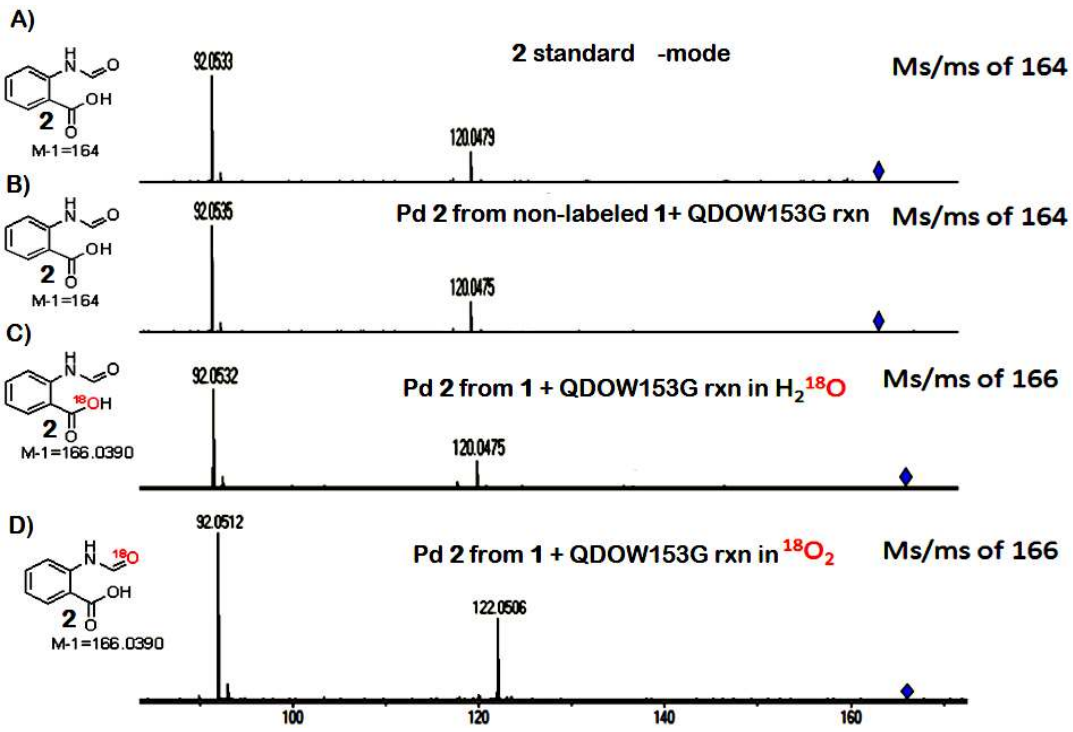
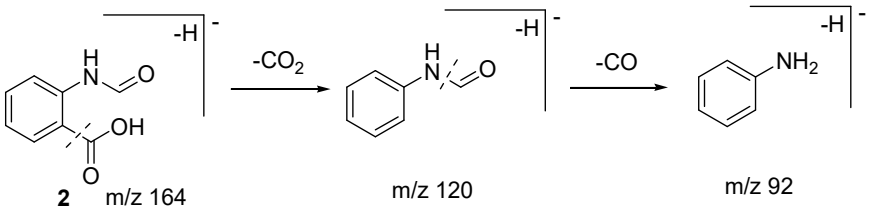
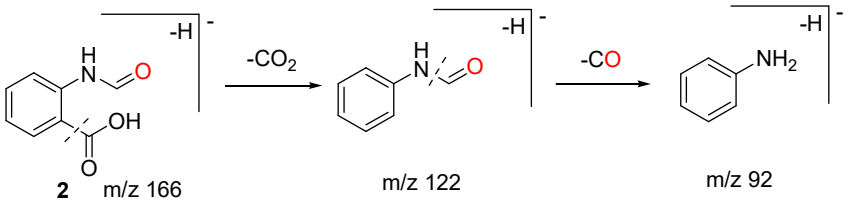


Figure 2.20: MS/MS data (in negative mode) of **2** generated from A) the synthetic standard (the MS/MS patterns maintained the same when dissolving the **2** standard in both $H_2{}^{16}O$ and $H_2{}^{18}O$ for 2 days); B) **1**+QDOW153G full reaction in non-label conditions; C) **1**+QDOW153G full reaction in $^{18}O_2$ (^{18}O is in red); D) **1**+QDOW153G full reaction in $H_2{}^{18}O$. (See analysis in Figure 2.20)

A) The MS/MS fragmentation of **2** standard and the product **2** from non-labeled **1**+QDOW153G full reaction



B) The MS/MS fragmentation of product **2** from **1**+QDOW153G full reaction in $^{18}O_2$



C) The MS/MS fragmentation of product **2** from **1**+QDOW153G full reaction in $H_2{}^{18}O$

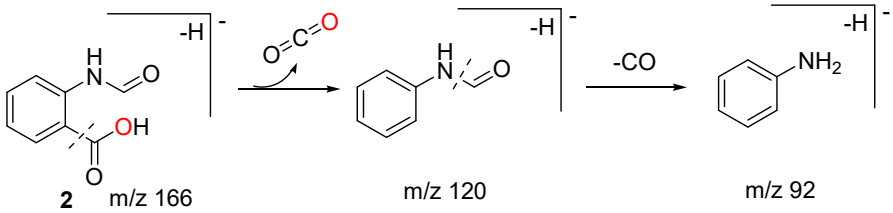


Figure 2.21: MS/MS analysis summary of **2**. A) The MS/MS fragmentation of **2** standard in both $H_2{}^{16}O$ and $H_2{}^{18}O$ for 2 days and the product **2** from non-labeled **1**+QDOW153G full reaction. B) The MS/MS fragmentatation of product **2** from **1**+QDOW153G full reaction in $^{18}O_2$. C) The MS/MS fragmentation of product **2** from **1**+QDOW153G full reaction in $H_2{}^{18}O$.

Besides, QDOW153G+**50** full reaction converted all **50** into the keto acid shunt product **52**, and almost no major CO-releasing product **54** was found in LCMS analysis (Figure 2.22). While in the QDOwt+**50** full reaction, the mass intensity ratio of **52** and **54** formation was around 1:10. This result further confirmed that the significant open-up of the QDOW153G binding pocket largely promoted the solvent trapping of the first-generated acylium ions **51** before any CO releasing (to form the second acylium ion **53**). Thus, barely any major product **54**

could be observed. In line with this, QDOW153G+**1** full reaction also had the accumulation of its keto acid shunt product **56** eluted at 6 min with $[M-H]^-$=192.0 Da (Figure 2.23) which was hardly detected in QDOwt+**1** reaction.

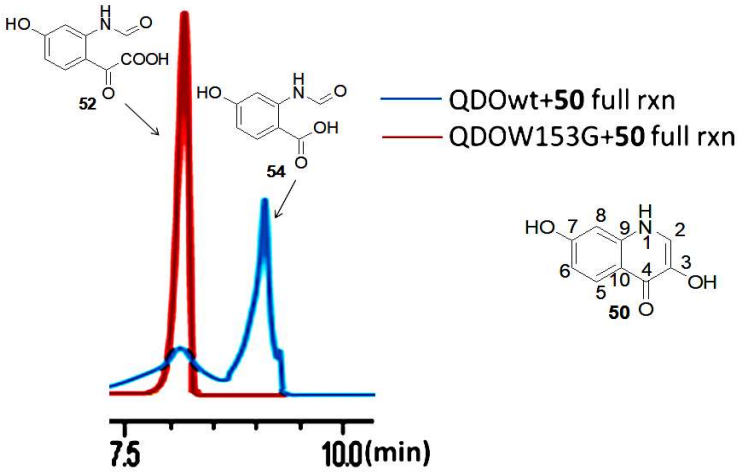


Figure 2.22: HPLC analysis of QDOW153G+**50** full reaction in comparison with QDOwt+**50** full reaction at 254nm. QDO W153G converted almost all **50** to the shunt product **52** at 8min and barely any major product **54** could be observed (red trace). The peak eluted at 9.2min in the blue trace is the major product **54** in wild type QDO full reaction with **50**.

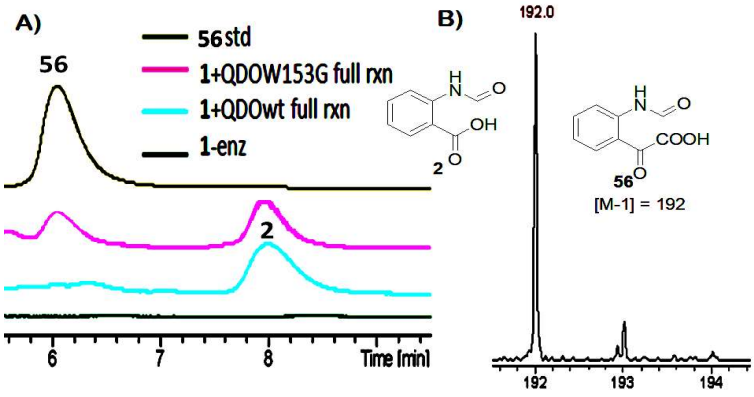


Figure 2.23: Accumulation of **56** in **1**+QDOW153G full reaction. A) HPLC of **1**+QDOW153G full reaction major product **2** at 8min and shunt product **56** at 6min. **1**+QDOwt full reaction did not show the obvious formation of **56**; B) MS of **56** from **1**+QDOW153G full reaction.

2.7 Conclusion

We have provided strong experimental evidence to demonstrate that QDO has followed the acylium ion mechanism proposed in Figure 10B. By engineering either the substrate or the QDO enzyme, the two acylium ion active species **6** and **7** (See Figure 2.8) successfully were trapped by the solvent water in the $^{18}O_2$ labeling reaction. To our knowledge, it is the first case of enzyme-catalyzed acylium ion formation being reported. Similar mechanisms involved in stable aromatic acylium ion intermediates may also apply to HOD (QDO's homolog), Quercetin 2,3-dioxygenase (2,3-QD), Acireductone dioxygenase (E-2)[47] and 2,4'-Dihydroxyacetophenone dioxygenase (DAD)[43]. High-resolution LC-MS played a key role in detecting the trace amount of shunt product **52** formation. The high sensitivity and accuracy of this technique may allow general applications in trapping and identifying tiny quantities of the mechanism-informative intermediates 'leaking' from the active site in various enzymatic systems.

2.8 Materials and methods

Materials: All chemicals were purchased from Sigma-Aldrich unless specified otherwise. LB-Lennox broth was from EMD Millipore. Kanamycin and IPTG were obtained from Lab Scientific Inc. Chloramphenicol was from Fisher Scientific. Primers for cloning and mutagenesis were obtained from Integrated DNA Technologies (IDT). The QIAprep Spin Miniprep Kit was obtained from QIAGEN. The Quikchange site-directed mutagenesis kit was obtained from Agilent Technologies. The Histrap column was obtained from GE Healthcare. Econo-Pack 10DG desalting columns were purchased from Bio-Rad. D_2O, D-chloroform, D6-DMSO, D4-methanol, and $H_2{}^{18}O$ were purchased from Cambridge Isotope Laboratories. 2.5L

baffled ultra yield flasks for overexpressing proteins were from Thomson Instrument Company. Qdo was gene synthesized and cloned into a pET28b vector by GenScript.

Overexpression and purification of QDO: Gene of Qdo was cloned in pET28b vector and $CaCl_2$-transformed into *E. coli* BL21(DE3) competent cells. 15ml of starter culture was grown in LB medium containing 50 μg/ml kanamycin overnight at 37°C. It was then added to 1.5L LB/kanamycin media as the seed. The flask with 1.5L medium was grown at 37°C with 180 rpm shaking until OD600 around 0.6. The culture was then induced by 1M IPTG and incubated at 15°C for 16 hr with 180 rpm shaking. The cells were collected by centrifugation (5000 rpm, 20 min) and stored at -80°C until further use. Typical cell yield was around 5 g from 1.5 L culture.

To purify the protein, the cell pellets from 1-1.5 L culture were mixed with 20-30 ml lysis buffer (50 mM KPi, 300 mM NaCl, pH 7.5) and thawed on ice with an addition of 6-8 mg lysozyme for 1 hr. The cell suspension was stirred to homogeneous and sonicated (the Misonix Sonicator 3000 model, 1.5s on/1.5s off pulses for 30s intervals, 5-6 times, allowing 5 min cooling time on ice between each interval) to lyse the cells thoroughly. The cell debris was removed by centrifugation at 15000 rpm for 30 min. The supernatant cell lysate was filtered through 0.45 μm filters and loaded onto a Ni-NTA His-trap column (equilibrated with lysis buffer). The loaded column was washed with 3-5 column volumes of wash buffer (lysis buffer + 50 mM imidazole) until all non-His-tagged proteins were washed out. The target protein was then eluted with the elution buffer (lysis buffer + 200 mM imidazole). The protein elute was concentrated to around 5 ml using 15ml 10kDa filters. The concentrated protein was passed through the Econo-Pac 10DG desalting column to remove imidazole. The desalted QDO protein was stored in the pH 7.5 PBS buffer (50 mM KPi, 300 mM NaCl) with 10% glycerol at -80°C.

SDS-PAGE gel showed a major purified protein band at 30kDa, which agreed with the molecular weight of QDO (Figure 2.24). The protein concentration could be obtained by Bradford assay, or by measuring the 280 nm absorbance (A280) and using the Beer's law to convert the absorbance into the concentration. The extinction coefficient could be obtained from the literature, or using ProtParam tool from the ExPASy proteomics server. The typical protein yield was 1mM*5ml (500μg) from 1.5 L cell culture.

Reconstitution of QDO activity: 200 μM QDO was incubated with 1mM substrate in pH7 PBS buffer in a final volume of 50 μl at room temperature for 30 min. The protein was denatured by an addition of 30 μl methanol and filtered through a 10 kDa 1.5 ml centrifugal filter. The filtrate was directly analyzed by HPLC or high resolution LC-MS with the injection volume of 25 μl. The substrate peak disappeared at 17 min and a new peak was found only in the full reaction mixture at 18 min. The product standard was synthesized and it could coelute with the new peak. LC-MS also showed the corresponding accurate mass of the product formation after the QDO catalyzed full reaction (Figure 2.24).

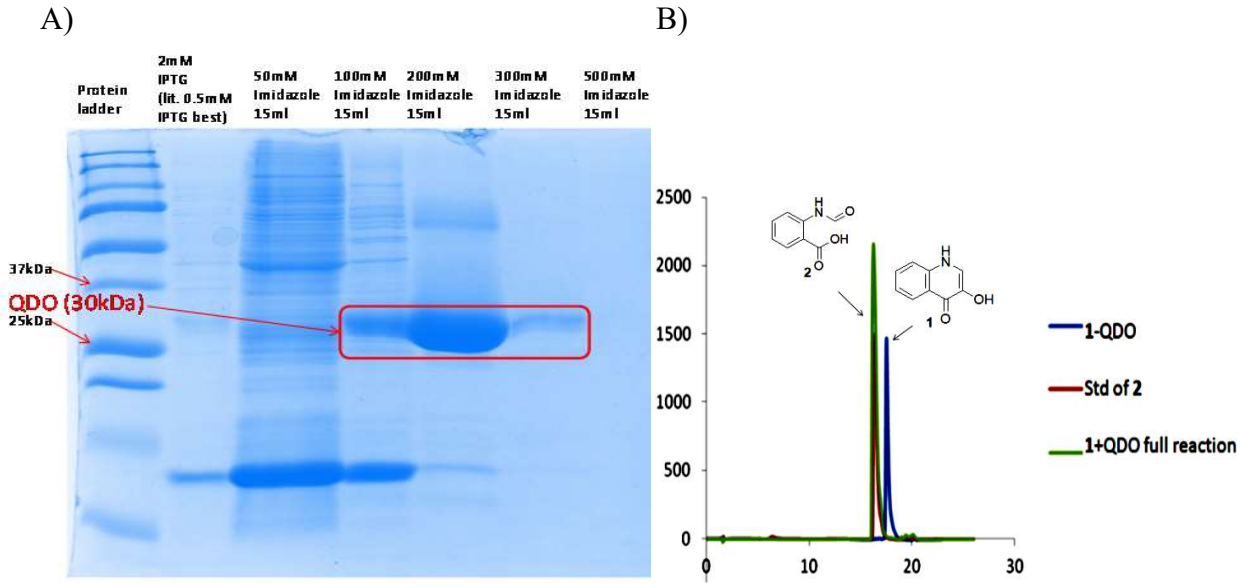


Figure 2.24: A) QDO overexpression; B) Activity reconstitution of **1**+QDOwt full reaction (determined the UV-vis absorption at 254nm)

HPLC parameters: The HPLC model was an Agilent 1200 or 1260 with a quaternary pump. The system included a diode array UV-Vis detector and LC separated components were detected using absorbance at 254, 280, 260, 295, 220 and 340 nm. The analysis was performed on a ZORBAX Eclipse XDB-C18 column (4.6 x 150 mm, 5 µm particles, Agilent Technologies). Data were processed using ChemStation version B.04.01 SP1 (Agilent Technologies).

HPLC conditions:

A- water

B- 10mM Ammonium acetate, pH 6.6

C- Methanol

HPLC method: (Flow rate = 1 ml/min)

0 min: 100% A; 7 min: 90% A, 10% B; 12 min: 25% A, 60% B, 15% C; 16 min: 25% A, 10% B, 65% C; 18 min: 100% B; 26 min: 100% B.

LC-MS parameters: LC-ESI-TOF-MS was performed using an Agilent 1260 HPLC system equipped with a binary pump and a 1200 series diode array detector followed by a MicroToF-Q II mass spectrometer (Bruker Daltonics) using an ESI source either in positive mode or negative mode. The analysis was performed on a Poroshell 120 EC-C18 column (3.0 x 100 mm, 2.7 μm particles, Agilent Technologies) and a Luna® C5 100 Å, LC column (100 x 4.6 mm, 5 μm particles, Phenomenex Inc.).

LC conditions:

A: 5 mM ammonium acetate buffer, pH 6.6

B: 75% Methanol and 25% Water

LC method: (Flow rate = 0.4 ml/min, for both the positive and negative mode on MS)

0 min: 100% A, 0% B; 7 min: 100% A, 0% B; 12 min: 80% A, 20% B; 16 min: 13.3% A, 86.7% B; 18 min: 100% A, 0% B; 26 min: 100% A, 0% B.

^{18}O labeling reaction in $H_2{}^{18}O$: 1 μl of QDO stock (1 mM in PBS buffer, pH 7.0) was added to 25 μl 1 mM **1** or **8** solutions in $H_2{}^{18}O$. The reaction mixture was incubated in the air at room temperature for 30 min. Methanol was added to quench the reaction. A 1.5ml 10K NMWL centrifugal filter unit was used to remove the protein. The filtrate was directly analyzed by LC-MS.

^{18}O labeling reaction in $^{18}O_2$: The QDO substrate **1** or **8** was dissolved in 50 μl PBS buffer (pH 7.0) to make a 1mM solution in a sealed Eppendorf tube and was purged with Ar for 20 min to remove the air ($^{16}O_2$). The tube was then connected to a balloon of $^{18}O_2$. 2 μl of QDO stock (1 mM in PBS buffer, pH 7.0) was added into the substrate solution that was exposed to $^{18}O_2$ and was incubated for 30 min at room temperature. The reaction was quenched with

methanol. The reaction mixture was passed through a 1.5ml 10K NMWL centrifugal filter unit and analyzed by LC-MS.

Site-direct mutations: Site-direct mutants were done using the QuikChange II Site-Directed Mutagenesis Kit from Agilent Technologies. The wild type Qdo gene template was as follows (optimized by Genscript):

ATGCAAAGCCTGAATGTGAACGGTACCCTGATGACCTACTCTGAATCTGGCGATCCG
CACGCACCGACCCTGTTCCTGCTGTCTGGTTGGTGCCAGGATCATCGTCTGTTTAAA
AACCTGGCACCGCTGCTGGCCCGTGACTTCCATGTGATTTGTCCGGATTGGCGCGGC
CACGACGCAAAGCAGACCGATAGCGGTGATTTTGACTCGCAGACGCTGGCCCAAGA
CCTGCTGGCATTTATTGATGCTAAAGGCATCCGTGACTTCCAGATGGTTAGCACCTC
TCACGGTTGCTGGGTGAATATCGATGTTTGTGAACAACTGGGTGCAGCACGTCTGCC
GAAGACGATTGTCATCGATTGGCTGCTGCAGCCGCATCCGGGCTTTTGGCAGCAACT
GGCGGAGGGTCAGCACCCGACCGAATATGTCGCCGGTCGTCAATCCTTTTTCGATGA
ATGGGCGGAAACCACGGATAACGCCGACGTGCTGAACCATCTGCGCAATGAAATGC
CGTGGTTCCACGGTGAAATGTGGCAGCGTGCATGCCGCGAAATTGAAGCTAATTACC
GTACGTGGGGCAGCCCGCTGGATCGCATGGAATCTCTGCCGCAGAAACCGGAAATT
TGTCATATCTATAGTCAGCCGCTGTCCCAAGATTACCGCCAGCTGCAACTGGACTTT
GCAGCTGGCCATAGCTGGTTCCATCCGCGTCACATCCCGGGTCGCACCCACTTCCCG
TCCCTGGAAAACCCGGTCGCCGTCGCACAAGCTATCCGTGAATTTCTGCAAGCGTGA

The primers (5’ to 3’) are:

H96G: CCAGATGGTTAGCACCTCTGGTGGTTGCTGGGTGAATATCG

H244G: CATCCCGGGTCGCACCGGTTTCCCGTCCCTGGAAAAC

S95G: CCAGATGGTTAGCACCGGTCACGGTTGCTGGGTG

Q32A: GCTGTCTGGTTGGTGCGGTGATCATCGTCTGTTTAAAAACC

S95G&H96G: CCAGATGGTTAGCACCGGTGGTGGTTGCTGGGTGAATATCG

W153G: CAATCCTTTTTCGATGAAGGTGCGGAAACCACGGATAACG

W153A: CAATCCTTTTTCGATGAAGCGGCGGAAACCACGGATAACG

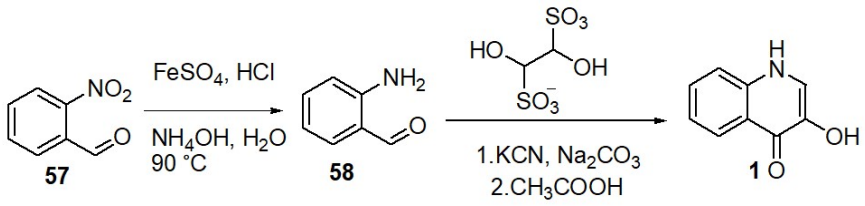


Figure 2.25: Synthesis of 1-H-3-hydroxy-4-oxoquinoline (**1**)

Synthesis of o-Aminobenzaldehyde (58)[48]**:** o-Aminobenzaldehyde (**58**) was obtained by the reduction of o-nitrobenzaldehyde (**57**) (Figure 2.25). The reaction mixture contained 1.2 g o-nitrobenzaldehyde, 21.0 g Iron(II) sulfate heptahydrate, 35 ml distilled water and 0.1 ml concentrated hydrochloric acid. The mixture was heated to a temperature of ~90 °C. Then 5 ml of concentrated ammonium hydroxide (NH_4OH) was added, after which, 6 ml ammonium hydroxide was slowly added in a two min interval. After 10 min of reaction, 10 ml×2 ether was used to extract the product from the reaction mixture. The organic layer was dried over anhydrous sodium sulfate and the ether solvent was removed under the vacuum condition using a rotary evaporator. The colorless product (**58**) crystals were obtained with the yield ~77%. The freshly made **58** was continually used for **1** synthesis. ^{1}H (300 MHz, DMSO-D_6): δ 7.11 (br s, 2H), 6.65 (d, 1H, J = 8.0 Hz), 6.73–6.77 (m, 1H), 7.28–7.34 (m, 1H), 7.54 (d, 1H, J = 8.0 Hz)

9.82 (s, 1H); 13C (300 MHz, DMSO-D_6): δ 115.9, 116.4, 118.8, 135.2, 135.7, 149.8, 194.1; ESI-MS (m/z): 122.1 (M+1).

Synthesis of 1-H-3-hydroxy-4-oxoquinoline (1)[49]: 0.7 g potassium cyanide was added to a stirred aqueous solution of sodium carbonate (20 ml, l M) under argon protection. 1.9 g glyoxal bisulphite was added to the mixture followed by immediate addition of 0.6 g o-aminobenzaldehyde (**58**). A small portion of THF was added to increase the solubility of o-aminobenzaldehyde. The solution became dark green and after 10-15 min a colorless precipitate was deposited. After a further 30 min, glacial acetic acid was added to adjust the reaction mixture pH to 6. The mixture was further stirred for 3 h. The precipitate was collected, washed with cold water, dried and recrystallized from ethanol to give 1-H-3-hydroxy-4-oxoquinoline (**1**) as a white powder (0.3 g, 30% yield) (Figure 2.25 and 2.26). ^{1}H (300 MHz, DMSO-D_6): δ 7.17-7.23 (m, 1H), 7.46-7.52 (m, 1H), 7.60 (d, 1H, *J* = 8.2 Hz), 7.77 (s, 1H), 8.11 (d, 1H, *J* = 7.89 Hz); ^{13}C (300 MHz, DMSO-D_6): δ 118.5, 122.3, 122.6, 124.2, 126.8, 131.5, 137.7, 138.5, 177.7; ESI-MS (m/z): 162.1 (M+1).

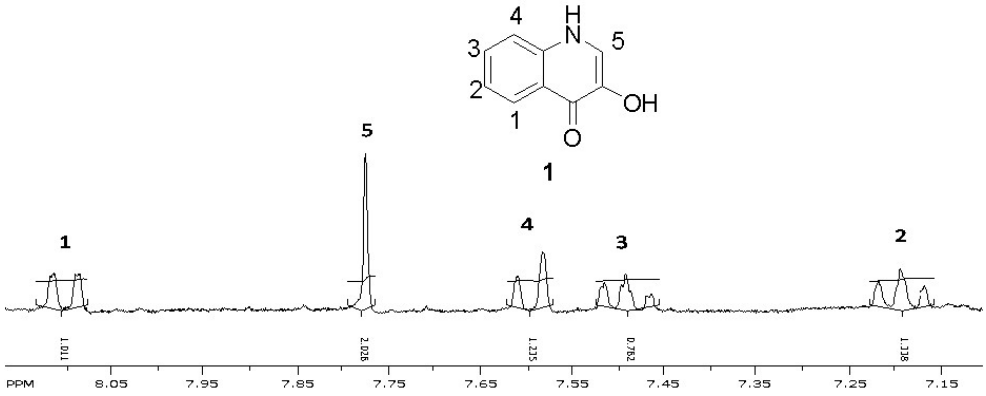


Figure 2.26: ^{1}H NMR (300 MHz, DMSO-d_6) of compound **1**

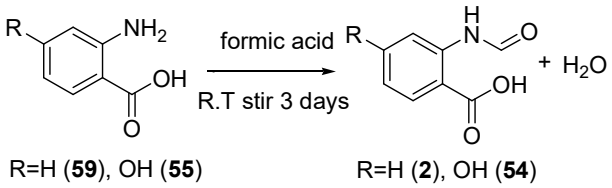


Figure 2.27: Synthesis of 2-formamidobenzoic acid (**2**). A similar route could also be used to synthesize 2-(2-formamidophenyl)-2-oxoacetic acid (**54**) using 2-amino-4-hydroxybenzoic acid (**55**) as the starting material.

Synthesis of 2-formamidobenzoic acid (2)[50] **and 2-(2-formamidophenyl)-2-oxoacetic acid (54):** 10 mmol anthranilic acid (**59**) or 2-amino-4-hydroxybenzoic acid (**55**) was dissolved in 5 ml 98-100% formic acid and was stirred for three days at room temperature. The white precipitate generated was collected and washed with a small quantity of formic acid. Then the crude product was air dried and recrystallized from ethyl acetate. The product white crystals had a yield of ~70%. 2-formamidobenzoic acid (**2**) (Figure 2.27-2.29): ^{1}H (300 MHz, DMSO-D_6): δ 7.12-7.21 (m, 1H), 7.52-7.62 (m, 1H), 7.98 (d, 1H, J = 8.1 Hz), 8.50 (s, 1H), 8.52 (d, 1H, J = 7.9 Hz), 11.18 (s, 1H); ^{13}C (300 MHz, DMSO-D_6): 114.5, 117.7, 121.2, 131.5, 131.9, 141.1, 160.5, 170.1; ESI-MS (m/z): 164.0 (M-1). 2-(2-formamidophenyl)-2-oxoacetic acid (**54**): ^{1}H (300 MHz, DMSO-D_6): δ 6.32 (d, 1H, J = 2.6 Hz), 6.37 (dd, 1H, J = 7.62, 2.56 Hz), 7.34 (s, 1H), 7.71 (d, 1H, J = 7.62); ^{13}C (300 MHz, DMSO-D_6): 97.7, 114.3, 115.0, 127.8, 141.1, 155.3, 162.5, 173.4; ESI-MS (m/z): 180.0 (M-1).

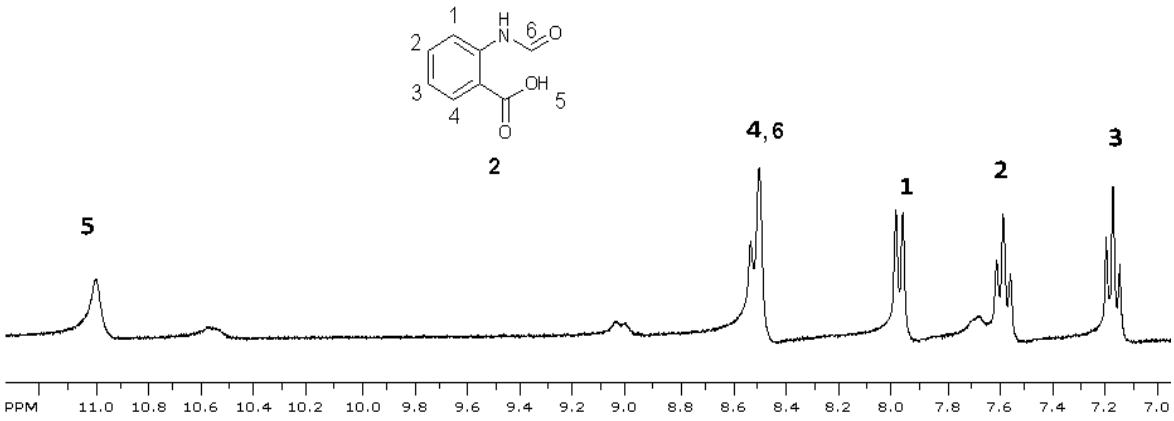


Figure 2.28: ^{1}H NMR (300 MHz, DMSO-d_6) of compound **2**

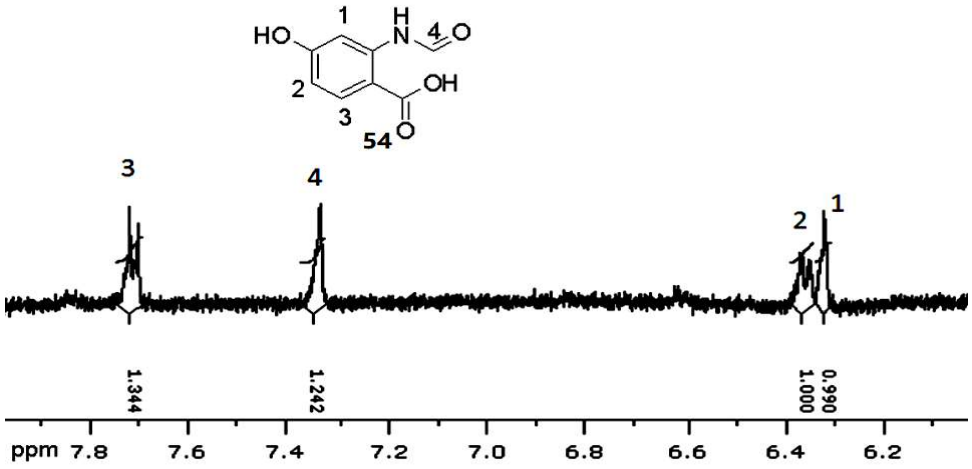


Figure 2.29: ^{1}H NMR (300 MHz, D_2O) of compound **54**

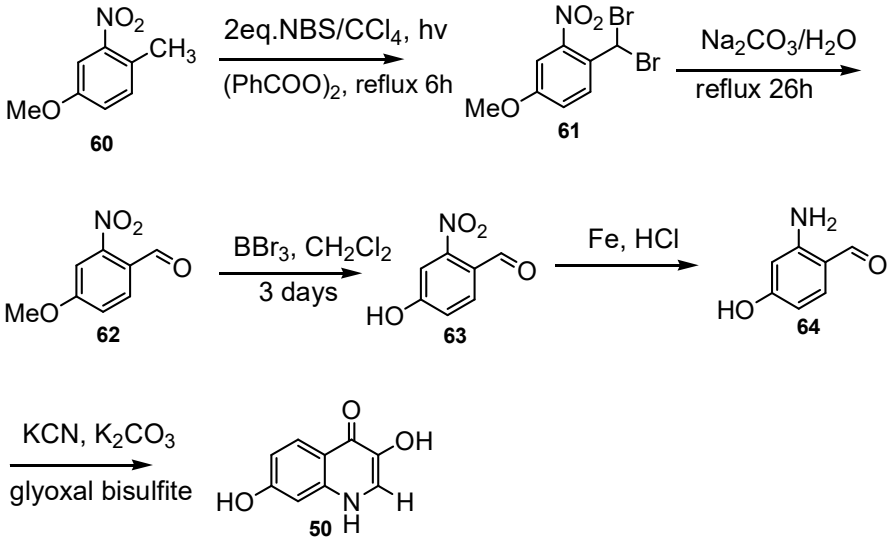


Figure 2.30: Synthetic route of 3,7-dihydroxyquinolin-4(1H)-one (**50**)

Synthesis of 1-dibromomethyl-4-methoxy-2-nitrobenzene (61)[51] **(Figure 2.30):** Benzoyl peroxide (0.40 g, 10 %) and N-bromosuccinimide (12.8 g, 72 mmol) were added in portions to a solution of 4-methyl-2-nitroanisole (**60**) (4.00 g, 24 mmol) in CCl_4 (30 ml). The reaction mixture was refluxed at 85 °C under an argon atmosphere for 12 h. The reaction mixture was cooled and the white precipitate (solid succinimide) was removed through filtration. The filtrate was concentrated to produce a crude product. The latter was recrystallized with dichloromethane-hexane to get a colorless crystalline solid product. The yield was 7.22 g, 95 %. ^{1}H (300 MHz, DMSO-D_6): δ 7.08 (dd, 1H, J = 8.3, 1.4 Hz), 7.21 (d, 1H, J = 8.3 Hz), 7.63 (d, 1H, J = 1.4 Hz), 6.66 (s, 1H), 3.73 (s, 3H); ^{13}C (300 MHz, DMSO-D_6): 37.5, 55.4, 100.9, 114.5, 127.8, 130.6, 148.3, 158.1; ESI-MS (m/z): 323.9 (M+1).

Synthesis of 4-methoxy-2-nitrobenzaldehyde (62)[52]**:** A mixture of 1-dibromomethyl-4-methoxy-2-nitrobenzene (**61**) (4.2 g, 24 mmol) and $NaHCO_3$ (6 g, ~72 mmol, 3 Eq) in water (40 mL) was refluxed for 26 h. The reaction mixture was cooled to room temperature and the

product was extracted with EtOAc (2x20 ml). The organic layer was washed with a solution of 1.5N HCl (2x10 mL), water (2x10 mL) and brine then dried over Na_2SO_4. The solvent was removed under reduced pressure to afford 1.9 g (80% yield) of the target compound **62** as a brown solid. ^{1}H (300 MHz, D_2O): δ 7.26 (dd, 1H, *J* = 8.0, 1.8 Hz), 7.55 (d, 1H, *J* = 1.8 Hz), 8.02 (d, 1H, *J* = 8.0 Hz), 10.33 (s, 1H), 1.64 (s, 3H); ^{13}C (300 MHz, D_2O): 50.4, 104.8, 112.6, 123.5, 130.4, 148.4, 162.2, 193.1; ESI-MS (m/z): 182.0 (M+1).

Synthesis of 4-hydroxy-2-nitrobenzaldehyde (63)[53]: 1g (5.5 mmol) 4-methoxy-2-nitrobenzaldehyde (**62**) was dissolved in 40 ml dichloromethane in a 100 ml round-bottom flask. 17.4 ml boron tribromide (1.0 M in dichloromethane, contained 17.4 mmol BBr_3) was added dropwise to the flask which was cooling in an ice bath. The reaction was further stirred in the ice bath for 30 minutes and then at ambient temperature for 3 hours. The reaction mixture was carefully poured into ice water and stood at ambient temperature for 3 days. The aqueous mixture was extracted with EtOAc, washed with sat. aq. NaCl, dried over Na_2SO_4, filtered and concentrated under reduced pressure. The crude material was purified by a SiO_2 gel flash chromatography using a linear gradient of 20 to 60% EtOAc in hexanes (collected the fraction when flushing with 50% EtOAc in hexanes), affording the product (**63**) as orange crystals (90% yield). ^{1}H (300 MHz, DMSO-D_6): δ 7.26 (dd, 1H, *J* = 7.9, 1.5 Hz), 7.41 (d, 1H, *J* = 1.5 Hz), 7.90 (d, 1H, *J* = 7.9 Hz), 10.03 (s, 1H); ^{13}C (300 MHz, DMSO-D_6): 100.9, 115.0, 127.8, 123.4, 143.3, 156.6, 197.3; ESI-MS (m/z): 168.0 (M+1).

Synthesis of 4-hydroxy-2-aminobenzaldehyde (64): A mixture of 0.3 g 4-hydroxy-2-nitrobenzaldehyde (**63**), 6 ml glacial acetic acid and 3 ml water was under an inert atmosphere of argon and then Fe powder (0.6 g) was added followed by 0.3 ml concentrated hydrochloric acid. The mixture was stirred at 60-65 °C for ~1 h (using TLC to

check the completion of the reaction) and cooled to room temperature. Then it was diluted with 14 ml water, neutralized with Na_2CO_3 to pH 6-7, and extracted with dichloromethane. The organic layer was washed with brine, dried over $MgSO_4$ and the CH_2Cl_2 was removed under reduced pressure to afford the product as an orange-brown solid (72% yield). ^{1}H (300 MHz, $CDCl_3$): δ 7.37 (d, 1H, *J* = 2.4 Hz), 7.47 (dd, 1H, *J* = 8.5, 2.4 Hz), 7.59 (d, 1H, *J* = 8.5 Hz), 9.51 (s, 1H); ^{13}C (300 MHz, $CDCl_3$): 99.4, 115.0, 128.4, 132.8, 147.3, 162.6, 199.2; ESI-MS (m/z): 138.1 (M+1).

Synthesis of 1-H-3,7-dihydroxy-4-oxoquinoline (50) (Figure 2.30): The reaction followed the procedure of 1-H-3-hydroxy-4-oxoquinoline (**1**) synthesis using 4-hydroxy-2-aminobenzaldehyde (**64**) as the starting material in replacement of o-aminobenzaldehyde (**58**). The crude product was recrystallized from ethanol-dichloromethane and further purified by prep-HPLC, as light brown powders (10-30% yield) (Figure 2.31). ^{1}H (300 MHz, CD_3OD): δ 6.80 (d, 1H, *J* = 1.9), 6.86 (dd, 1H, *J* = 8.8, 1.9 Hz), 8.11 (d, 1H, *J* = 8.8 Hz), 7.63 (s, 1H); ^{13}C (300 MHz, CD_3OD): 99.6, 118.0, 122.7, 124.2, 130.5, 140.7, 144.2, 155.3, 180.9; ESI-MS (m/z): 178.1 (M+1).

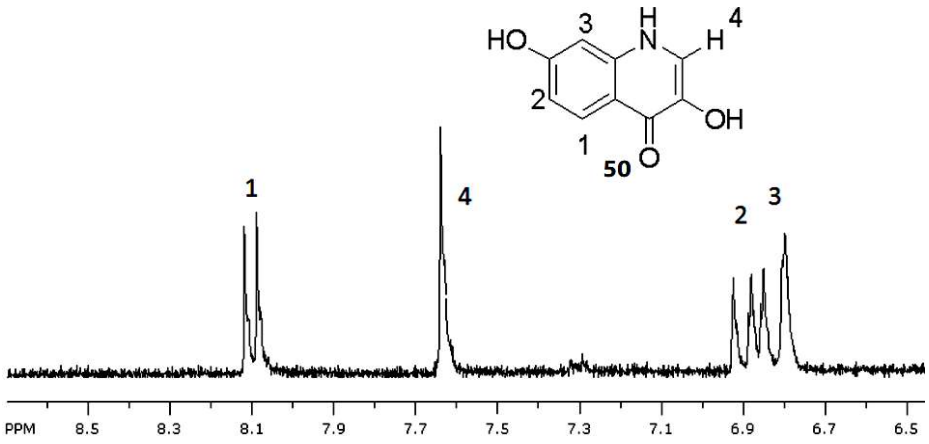


Figure 2.31: ^{1}H NMR (300 MHz, CD_3OD) of compound **50**

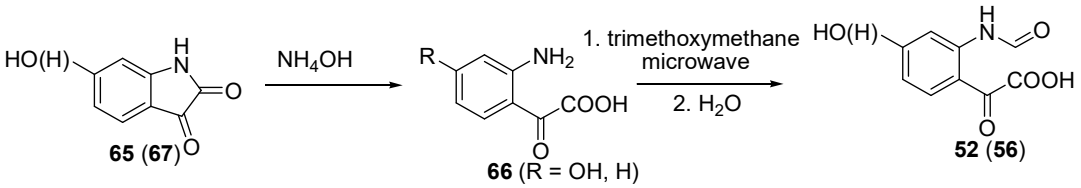


Figure 2.32: Synthesis of 2-(2-formamido-4-hydroxyphenyl)-2-oxoacetic acid (**52**). The similar route was used to synthesize 2-(2-formamidophenyl)-2-oxoacetic acid (**56**) with isatin (**67**) as the reactant instead of **65**.

Synthesis of 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid or 2-(2-aminophenyl)-2-oxoacetic acid (66): 0.2 g 6-Hydroxyisatin (**65**, from Santa Cruz Biotechnology) or isatin (**67**) was dissolved in 2 ml ammonium hydroxide solution (28.0-30.0% NH_3 basis) and stirred overnight at 30 °C. The solvent and the excess amount of NH_4OH were removed by a lyophilizer (Labconco FreeZone 4.5 Liter Freeze Dryer). The product yield was 90%. 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid: ^{1}H (300 MHz, DMSO-D_6): δ 6.28 (d, 1H, J = 1.6), 6.40 (dd, 1H, J = 7.8, 1.6 Hz), 7.37 (d, 1H, J = 7.8 Hz), 10.75 (s, 1H), 11.18 (s, 1H); ^{13}C (300 MHz, DMSO-D_6): 99.4, 115.0, 122.5, 124.4, 139.8, 156.6, 164.7, 186.1; ESI-MS (m/z): 182.0 (M+1). 2-(2-aminophenyl)-2-oxoacetic acid: ^{1}H (300 MHz, DMSO-D_6): δ 7.20-7.32 (m, 1H), 7.72-7.82 (m, 1H), 8.08 (d, 1H, J = 8.1 Hz), 8.33 (s, 1H), 8.52 (d, 1H, J = 7.9 Hz), 11.35 (s, 1H); ^{13}C (300 MHz, DMSO-D_6): 116.9, 122.2, 122.5, 126.8, 130.8, 151.2, 172.6, 190.1; ESI-MS (m/z): 166.1 (M+1).

Synthesis of 2-(2-formamido-4-hydroxyphenyl)-2-oxoacetic acid (52) and 2-(2-formaamidophenyl)-2-oxoacetic acid (56) (Figure 2.32): 0.1 g 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid or 2-(2-aminophenyl)-2-oxoacetic acid (**66**) was dissolved in 1 ml of anhydrous trimethoxymethane and microwaved for 20 min at power 2 and another 20 min at power 3 in a

1000 watt microwave. 5 ml of water was added after the reaction mixture was cooled down. The solvent was dried by a lyophilizer or a speed vac. The crude product was purified by HPLC to get a brown powder. The yield was ~40%. 2-(2-formamido-4-hydroxyphenyl)-2-oxoacetic acid (**52**) (Figure 2.33-2.35): ^{1}H (400 MHz, DMSO-D_6): δ 7.20 (d, 1H, *J* = 2.9), 7.32 (dd, 1H, *J* = 11.2, 2.8 Hz), 8.02 (d, 1H, *J* = 11.4 Hz), 9.10 (s, 1H), 9.33 (s, 1H), 10.87 (s, 1H); ^{13}C (400 MHz, DMSO-D_6): 108.9., 120.0, 121.5, 130.0, 152.0, 155.6, 159.3, 162.8, 163.1; ESI-MS (m/z): 208.0 (M-1). 2-(2-formaamidophenyl)-2-oxoacetic acid (**56**): ^{1}H (400 MHz, DMSO-D_6): δ 7.16-7.23 (m, 1H), 7.54-7.62 (m, 1H), 8.10 (d, 1H, *J* = 10.1 Hz), 8.47 (s, 1H), 8.62 (d, 1H, *J* = 9.9 Hz), 9.03 (s, 1H); ^{13}C (400 MHz, DMSO-D_6): 117.6, 123.4, 124.8, 132.6, 135.5, 141.1, 160.4, 164.7, 186.1; ESI-MS (m/z): 192.0 (M-1).

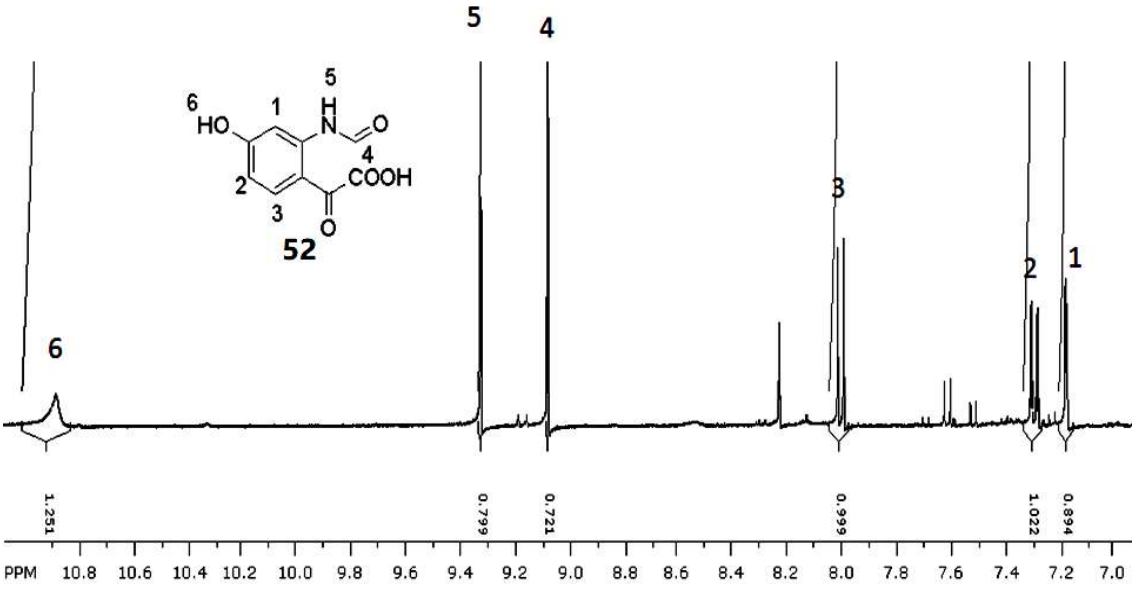


Figure 2.33: ^{1}H NMR (400 MHz, DMSO-d_6) of compound **52**

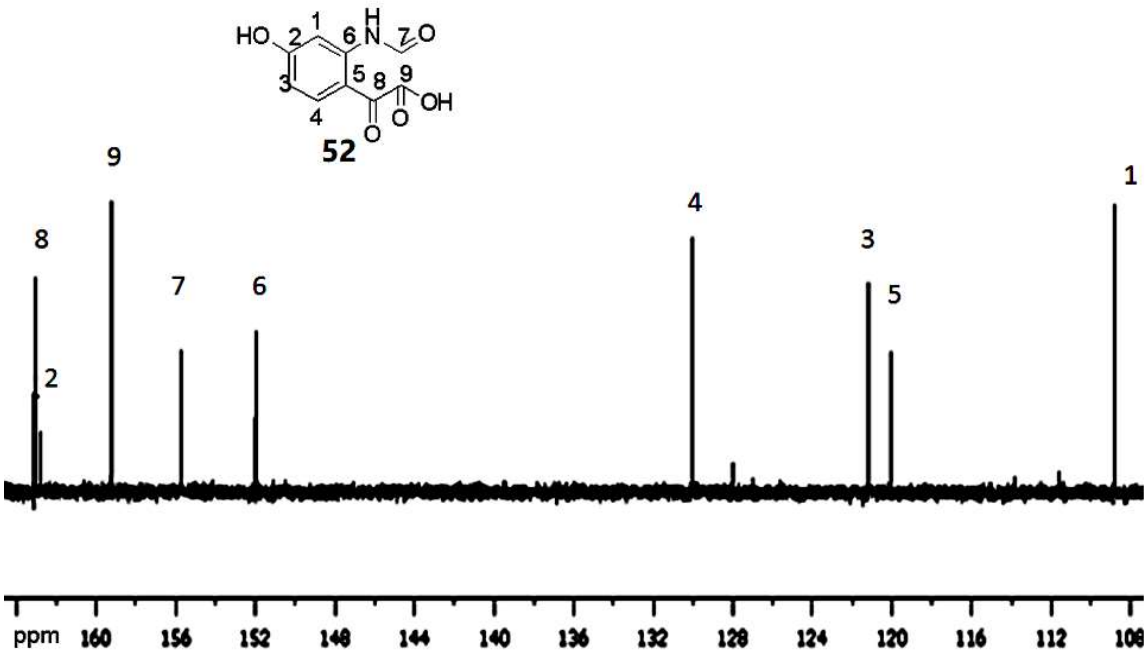


Figure 2.34: ^{13}C NMR (400 MHz, DMSO-d_6) of compound **52**

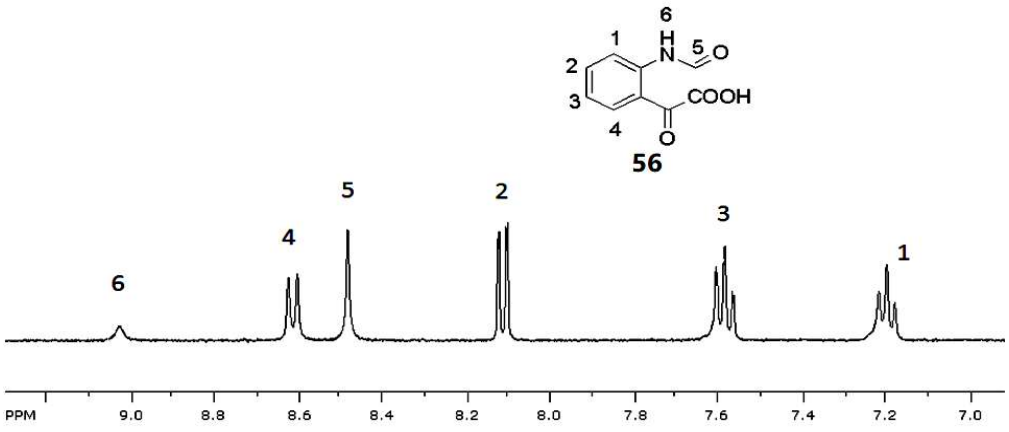


Figure 2.35: ^{1}H NMR (400 MHz, DMSO-d_6) of compound **56**

3. COMPUTATIONAL SUPPORTS ON QDO ACYLIUM ION MECHANISM

3.1 Computational chemistry

Computational chemistry is a sum of techniques that have been used on computers to solve chemistry problems. Scientists set mathematic models to describe the physical and chemical laws quantitatively. A good model not only rationalizes all the detectable facts but also can make predictions that lead to novel scientific discoveries and theoretical breakthroughs. The computer is the most powerful tool to deal with massive and complicated mathematical calculations. By combining computer science approaches (such as algorithm optimization, big data mining, and statistics) and theoretical chemistry principles, computational chemistry has become a distinct chemistry branch that has irreplaceable applications especially in drug design and new material development.

The critical feature of the computational chemistry is "approximation". Particles such as atoms and molecules are known to have wave-like properties and follow quantum mechanical rules described by the Schrödinger equation (Figure 3.1). The Schrödinger equation of a molecule defines the energy and distribution of its electrons, which are crucial properties in chemical reactions. However, the Schrödinger equation cannot be solved analytically for any molecule with more than one electron (e.g., H_2^+). Approximations must be used to simulate the many-body systems. The approximation strategies include neglecting some less contributed parts of the Schrödinger equation, adding empirical parameters and introducing the classical Newton's law to the model. Based on the different strategies of approximation, computational chemistry methods can be divided into five broad classes: Ab Initio, Density Function Theory, Semiempirical, molecular mechanics and molecular dynamic.[54]

$$\hat{H}\Psi(\vec{r},t) = \frac{-\hbar}{t}\frac{\partial\Psi(\vec{r},t)}{\partial t}$$

$\hat{H}$ Hamiltonian (operator)

Ψ Wavefunction (eigenfunction)

t time

$\hbar$ $\frac{h}{2\pi}$

h plank's constant ($6.62606957 \times 10^{-34}$ J · s)

Figure 3.1: Schrödinger equation

Ab Initio means “from the start” in Latin. In other words, it is “from theoretical principles”, but not from the experimental data. The fundamental theoretical principle of Ab Initio calculations is the Schrödinger equation. This method does not use any empirical or semiempirical parameters extracted from the experimental facts. All the approximations are based on quantum mechanical simulations. The outer layer of electrons are valued most in studying the chemical reactions. Other factors described in the Schrödinger equation can then be neglected to some extent, such as the time, the coupling of electronic and nuclear motion and the electron-electron repulsion. The various extent of approximation in the equation leads to different “levels of theory”. The Ab Initio calculation with the less approximation is considered to use the higher level of theory, which usually brings more accurate simulation results. However, the high level of the Ab Initio calculation is computationally costly and only suitable for the simple, small molecular systems. The chosen of the level of theory is dependent on the allowed range of error. Given the similar acceptable accuracy, the lower level of theory usually saves more computational time and resources.

Semiempirical is also based on the Schrödinger equation. To further save the computational cost, some very complicated calculations in the ab initio method are replaced by a library of parameters which are calibrated to fit the experimental data. This type of approximation is called "semiempirical" because it is a mixture of quantum theory and experimentally based parameters. It could be poor in accuracy if the parameters are improperly selected. However, it can be 100-1000 times faster than the ab initio method.

Density functional theory (DFT) is still based on the Schrödinger equation. Instead of calculating a conventional wavefunction as ab initio and semiempirical methods do, DFT methods derive the electron distribution directly from the electron density function. In a many-electron system, functionals—functions of another function are used to approximate the space-dependent electron density function. DFT methods can reach high accuracy that is comparable to the ab initio method with a much lower computational cost. Serious DFT calculations appeared in the 1980s, while the early ab initio and semiempirical calculations were achieved in the 1960s[55,56]. Walter Kohn and John Pople won the Chemistry Nobel Prize for their work on developing the density functional theory as a powerful computational tool in quantum chemical simulations in 1998. This relatively new method has become one of the most popular computational approaches involved in physics, chemistry, and material science applications. The DFT method is suitable to study the covalent bond cleavage and formation in a chemical reaction, but it has poor behaviors in calculating van der Waals forces, transition states, and global potential energy surfaces. Semiempirical and empirical corrections have been added to the DFT calculation to increase the simulation accuracy of such calculations[57].

Molecular mechanics (MM) is based on the classical spring model that follows Newton's law. Each atom can be considered as a ball with the assigned radius (usually the van

der Waals radius), polarizability and net charge derived from experiment and/or quantum calculations. The bonded interactions are like springs. Given the known spring lengths and angles between them, and the energy required to bend and stretch the springs, it is possible to figure out the exact geometry that a group of springs and balls can reach by spending the lowest amount of energy. In this way, the molecular geometry can be optimized without using quantum mechanical (QM) calculations. MM methods are heavily parameterized. They are much more computationally inexpensive when compared with the quantum calculations (e.g., ab initio, DFT and semiempirical methods). Though the electronic effect is ignored, MM calculations can still make good simulations in non-covalent interactions involved in the large biomolecule systems. Successful examples can be found in drug-enzyme binding geometry optimizations (docking), protein substrate binding constant calculations and binding site engineering, protein folding kinetics simulations, and protein-solvent interaction studies[58,59].

Molecular dynamics (MD) calculates the forces and potential energies of the system via molecular mechanics or quantum mechanics methods and then applies them to solve Newton's laws of motion to check the position and velocity changes of the particles involved in the system in a time-dependent way. Thus, one can simulate the motion of an enzyme as it changes shape on binding to a substrate or the motion of a swarm of water molecules around a molecule of solute. A trajectory of quantum mechanical molecular dynamics simulation is capable of modeling the real chemical reaction mechanism.[60]

To check the energy feasibility of our proposed acylium ion mechanism of QDO catalyzed reaction, we calculated the energy profiles of QDO+**1** reaction through both mechanisms (Figure 3.2) using the DFT methods. This work was in collaboration with Dr. Lisa M. Perez, the manager of the Laboratory for Molecular Simulation at Texas A&M University.

These simulations only contained minimal molecules that required for the reaction without the protein shell. With Lisa's training, I also used MM calculations on docking **1** into the QDO active site (QDO+**1** crystal is not available) and the geometry optimization of the entrance of the binding pocket of the W153G QDO enzyme.

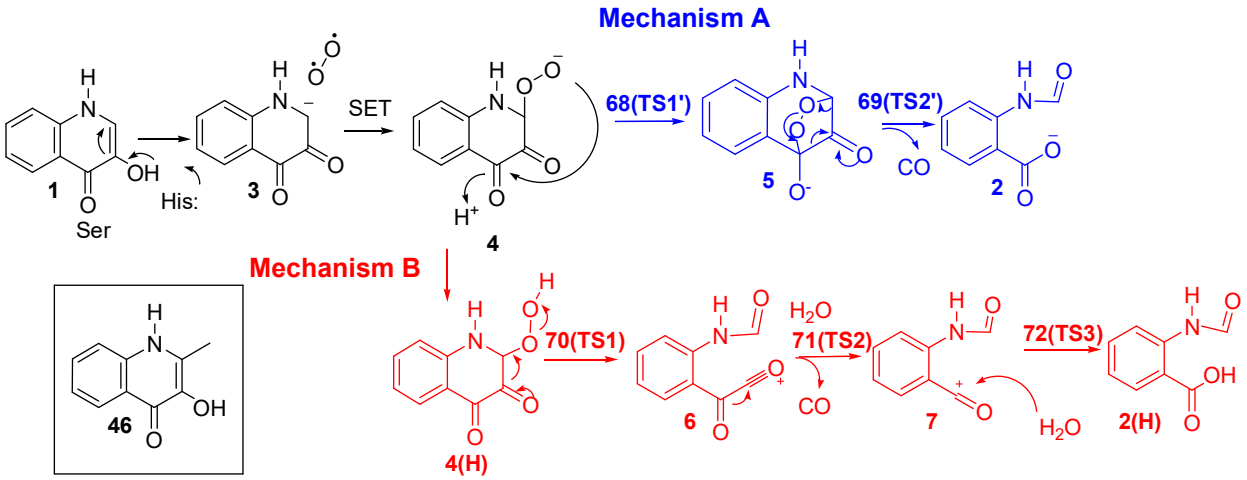


Figure 3.2: Alternative proposed mechanisms of QDO+**1** full. Mechanism A is the reported cyclic peroxide mechanism. Mechanism B is our proposed acylium ion mechanism. The optimized structures of transition states **68-72** are shown in Figure 3.3B. **46** is the native substrate of HOD, a homolog of QDO catalyzing the same type of reaction.

3.2 Results and discussions

3.2.1 DFT energy profile of both proposed mechanisms (Figure 3.2, Mechanism A & B) of QDO

The calculation of the initial O_2 addition to the HOD (homolog of QDO) substrate **46** (black-boxed in Figure 3.2) has been reported.[61] Our calculations mainly focused on the steps after the O_2 incorporation (Figure 3.3). The full reaction energy profile calculations for both mechanisms using substrate **1** were performed under implicit water solvation model with no protein shell. The reaction involved proton transfer, CO releasing, hydroxide releasing and

rebinding. To simplify and maintain the acceptable error range of the intermolecular reaction calculation, the energy of essential non-interacting species, O_2, 5-methylimidazole, acetic acid (mimic His and Asp as the proton acceptor/donor required in the reaction), CO and OH^-, were calculated separately and added together to balance the system charge and atoms. So the vast errors generated in the absolute free energy calculations were able to cancel out to obtain reliable relative-energy profiles (see Figure 3.3C for details of the add-up of each balanced species). The free energy profiles of both mechanisms were shown in Figure 3.3A. Compared with the starting substrate **1** in solution with no QDO enzyme, the reaction through both mechanisms required an activation barrier of ~40 kcal/mol. This result was in line with the experimental data that **1** required a strong base to form the product **2** in non-enzymatic aqueous solution.

A)

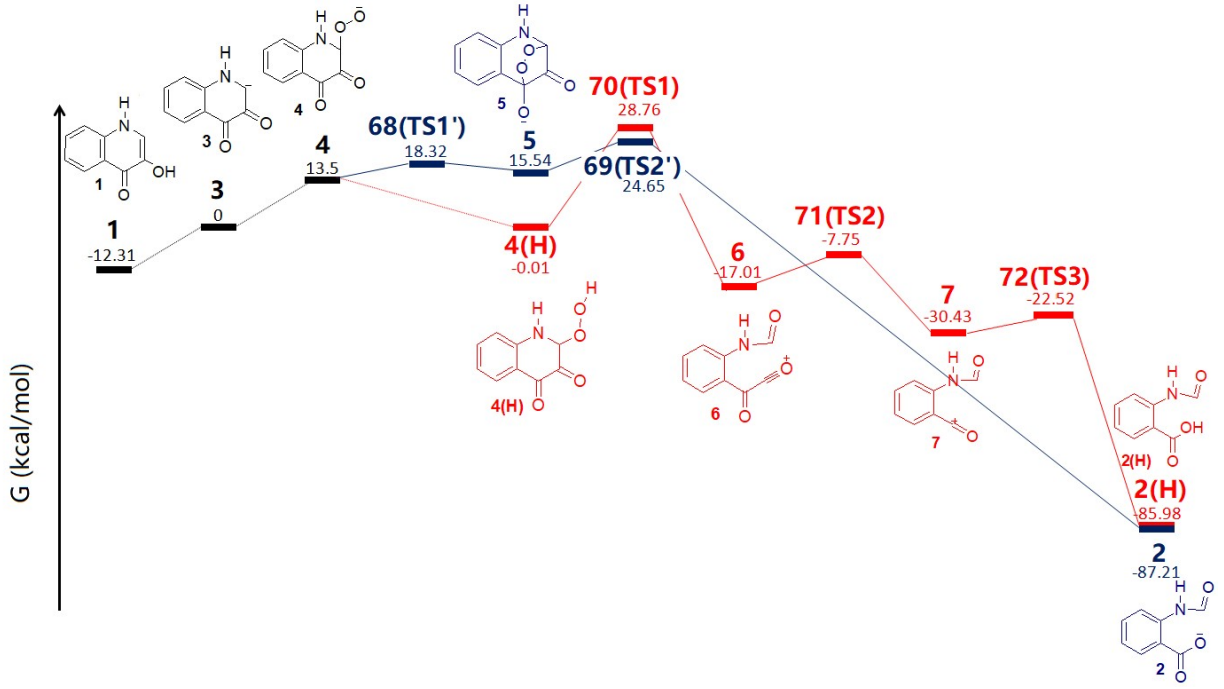
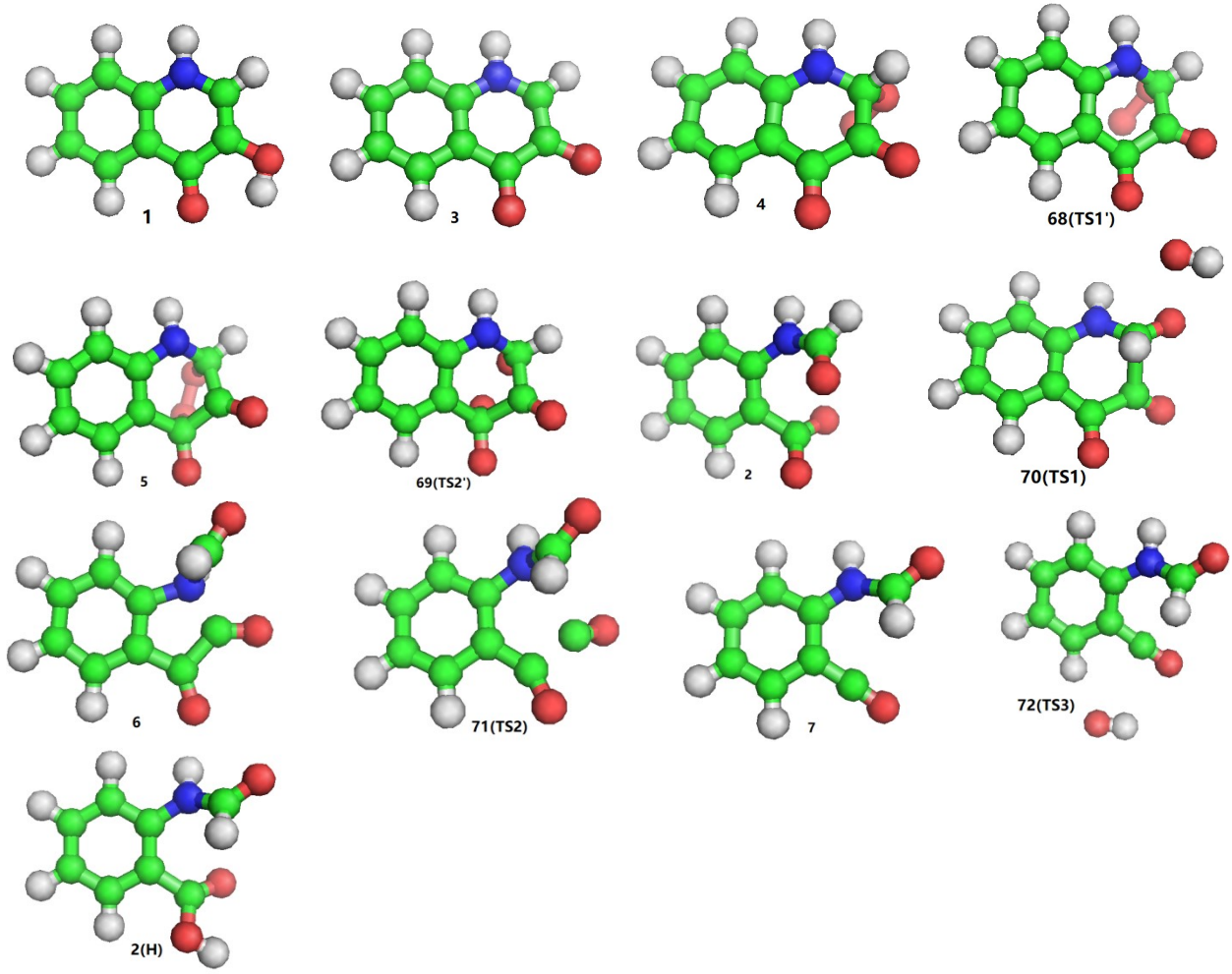


Figure 3.3: DFT energy profile diagrams for both the cyclic peroxide generated mechanism (in blue) and the acylium ion involved mechanism (in red). The common species of the two mechanisms in the initial steps are in black. The barrier between **4** to **4(H)** was calculated separately in Figure 3.4. The transition states between the common species **1** and **4** were not included in this calculation. B3LYP/6-311+G(d,p) (solvent: water). A) The DFT energy profiles of both mechanisms. B) The optimized structures of important intermediates and transition states. The carbon atom is in green, the hydrogen is in white, the oxygen is in red, and the nitrogen is in blue. C) Balanced species for the energy calculations.

B)



C)

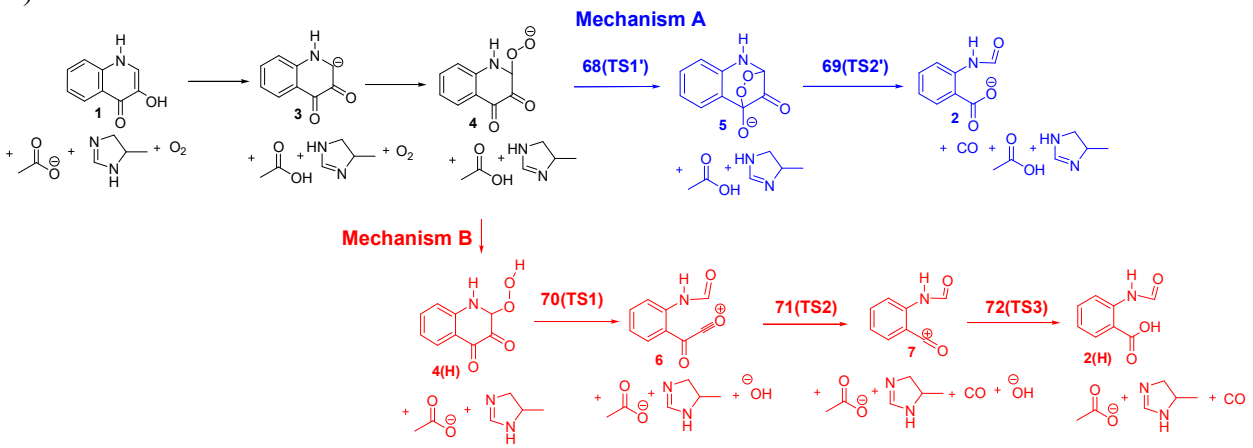


Figure 3.3: Continued

QDO was reported to catalyze the reaction as a general base. We set a bigger model including the important QDO binding pocket residues and the substrate **1** (Figure 3.4A) and found that the free energy of **3** was dramatically dropped to even 1.7 kcal/mol lower than **1** inside the QDO binding pocket (as compared with an uphill energy difference of 12.3 kcal/mol in the water salvation model, Figure 3.3A and 3.4B). The **1** to **3** conversion inside the QDO binding pocket only had a 0.5 kcal/mol barrier on the electronic energy (Ee) surface while the free energy calculation at 298 K suggested that it was a barrierless process (with the transition state energy lower than the starting material). Therefore, we set the relative free energy of **3** as the zero point to compare the energy barriers of both mechanisms.

DFT calculations (only included the substrate **1** and O_2 with implicit water solvent modification without the protein shell) on both mechanisms (Figure 3.3A) had an activation barrier of less than 30 kcal/mol in aqueous solution (28.8 kcal/mol for the acylium ion mechanism and 24.7 kcal/mol for the cyclic peroxide mechanism), which was reasonable for a reaction to happen at room temperature. The free energy barrier of acylium ion mechanism was ~4 kcal/mol higher than the reported cyclic peroxide mechanism in solution. The difference was within the acceptable error change (< 5 kcal/mol). Therefore, the QDO reaction could undergo either mechanism without significant energy preference. The reaction barrier could be much lower inside the QDO binding pocket. The acylium ion mechanism might be affected more by the protein shell protection than the cyclic peroxide mechanism, since cation intermediates were more sensitive to the solvent water approaching. When the QDO binding pocket was opened significantly through W153G mutation (W153 was reported having no catalytic but only structural effect on the QDO+**1** reaction), a much slower reaction rate was observed (>10^4 times slower than the wild-type QDO).[61]

The optimized geometry of intermediates and transition states were shown in Figure 3.3B. In particular, the first acylium ion **6** was a little bent rather than the ideal linear conformation due to the interaction with the lone pair electrons of the formyl amide nitrogen. The distance between the nitrogen and the acylium ion carbon was 1.6 Å, which was 0.1 Å longer than the typical C-N single bond length (1.5 Å). **6** also had a low CO releasing barrier of only 9 kcal/mol. Thus, the acylium ion carbon of **6** might form a dative bond with the nitrogen lone pair electrons which was weaker than a real covalent C-N bond but had some stabilization effect on this reactive keto acylium ion species. The formyl amide had to rotate for the better orbital overlapping with the acylium ion carbon (Figure 3.3B). When inside the QDO binding pocket, this rotation may be constrained by the protein shell, which may weaken the C-N dative bond and increase the CO releasing efficiency and the product specificity.

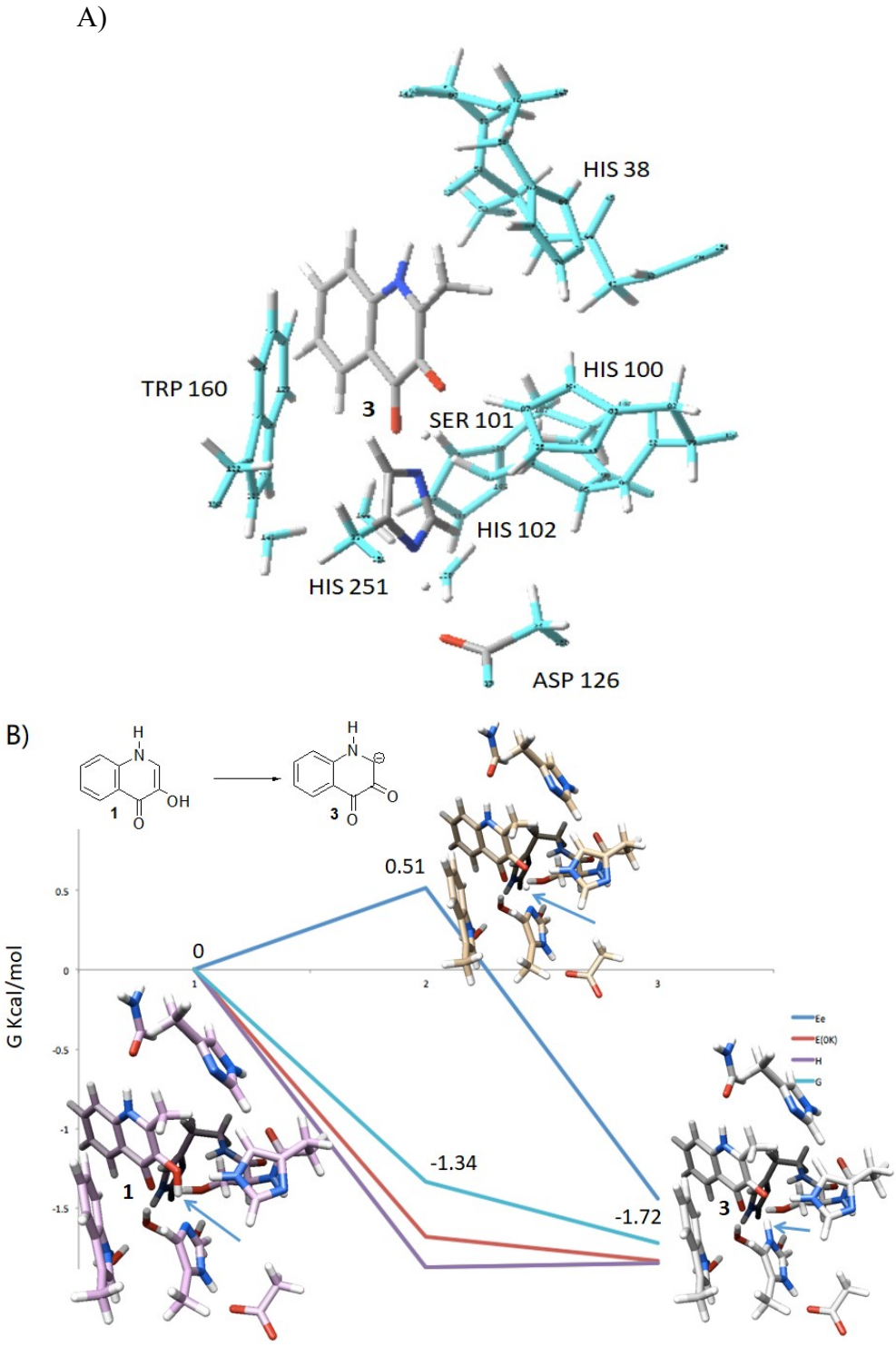


Figure 3.4: **1** to **3** calculations with important QDO binding pocket residues. A) Residues included in the model. The grey residues were allowed to move, while the cyan residues were frozen. B) The DFT energy profile and optimized geometries of **1** to **3** conversion. C) Overlapped structures of **1** (in green), **3** (in yellow), and the **1**-to-**3** transition state (in red).

C)

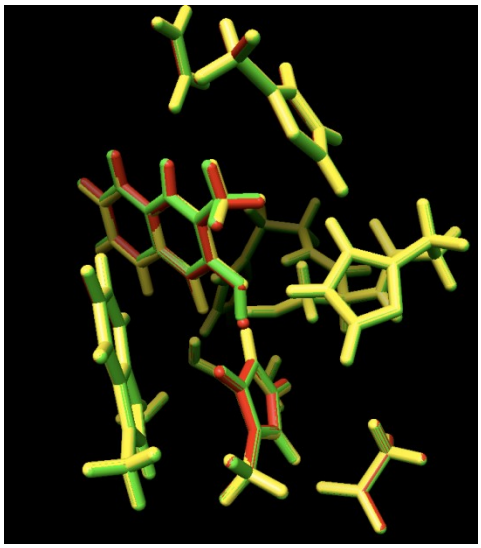


Figure 3.4: Continued

The peroxide anion **4** was protonated to form **4(H)** in the acylium ion mechanism. The crystal structural analysis of the HOD+**46** complex (PDB: 2WJ4) HOD was a homolog of QDO catalyzing the same type of reaction using the substrate 46, Figure 3.5) suggested that the proton source might be the protonated His251. His251 and Asp126 worked together as a general base which first took the proton from 3-OH of the substrate **46** to initiate the O_2 incorporation process and form **75**. Then the protonated His251 transferred the proton back to **75** to generate **75(H)**. A simple DFT modeling including **75**, a methanol (Ser101 mimic), a protonated 5-methylimidazole (His251 mimic) was performed to figure out the barrier of the proton transfer process (Figure 3.5). The highest energy barrier between **75** to **75(H)** was less than 4 kcal/mol. This protonation step could easily happen with such a low energy barrier.

The CO releasing barrier of **1** and **50** were also calculated to check whether the 7-OH substitution could stabilize or destabilize the first keto acylium ion (**6** and **51**, Figure 3.6). The barrier of 7-OH analog CO releasing was ~1 kcal/mol lower than the native substrate CO releasing barrier. Therefore **51** might be slightly easier to release CO than **6**. The accumulation of the side product **52** might be caused by the poor binding orientation of **50** inside the QDO binding pocket that decreased the product specificity.

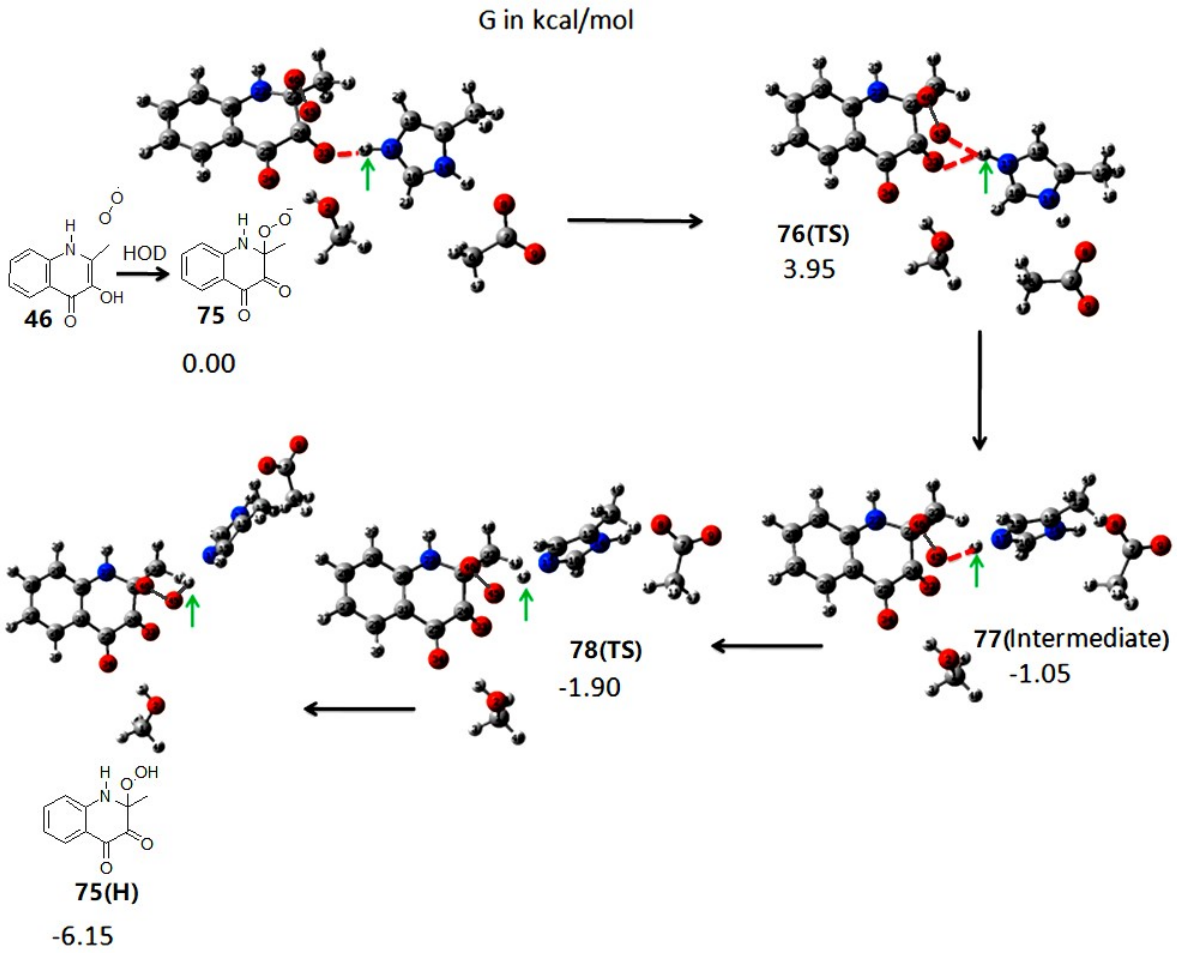


Figure 3.5: DFT energy diagram of the proton transfer from His251 side chain to the peroxide anion **75** to form **75(H)**. **75** is the analog of **4** generated by HOD using **46** as the substrate. The red dash line is the hydrogen bonding effect. The green arrow points out the transferred hydrogen

atom. The energy barrier is less than 4 kcal/mol. The carbon atom is in grey, the hydrogen is in white, the oxygen is in red, and the nitrogen is in blue.

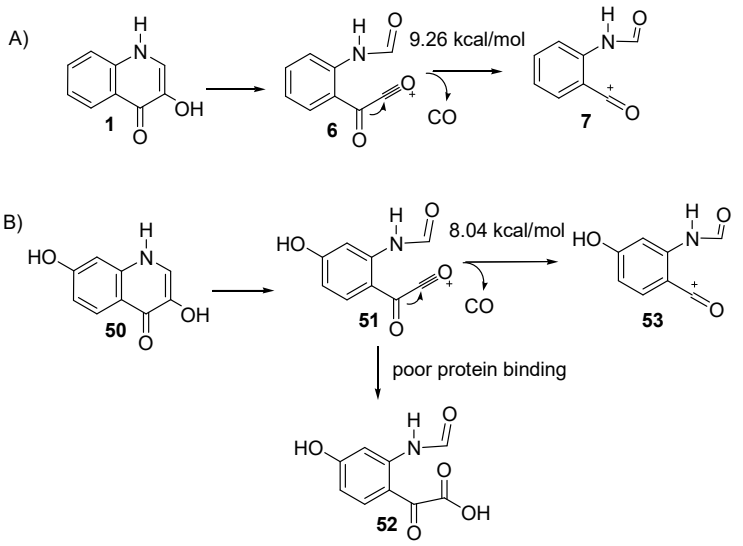


Figure 3.6: The The CO releasing step energy barrier comparison between A) the native substrate **1** derived acylium ion **6** (9.26 kcal/mol) and B) the 7-OH substituted substrate **50** derived acylium ion **51** (8.04 kcal/mol).

3.2.2 MM docking of 1 into QDO active site

QDO only has a protein structure with no substrate bound (3IBT), but its homolog HOD+**46** (**46** is the substrate of HOD) complex structure (2WJ4) has been reported[19]. Since the two protein structures are highly superimposed, the HOD+**46** structure is a good reference for docking **1** into the QDO (both the wild-type and the W153G mutant) active site (Figure 3.7). The HOD with or without substrate structures (2WJ4 and 2WJ3, respectively) do not have apparent conformational changes, which suggests that the protein structure is relatively rigid. Therefore, most QDO residues can be restrained during the docking process. Only the side chains of seven residues surrounding the active site (H244, H96, S95, W153, I185, Q32, and W30) were allowed to move when this docking was done. 81 confirmation matches (poses) were generated, and 5 of

them were seated inside the binding pocket with the reasonable orientations. The one with the lowest energy superimposed most to the HOD+**46** crystal structure (Figure 3.8).

A) B)

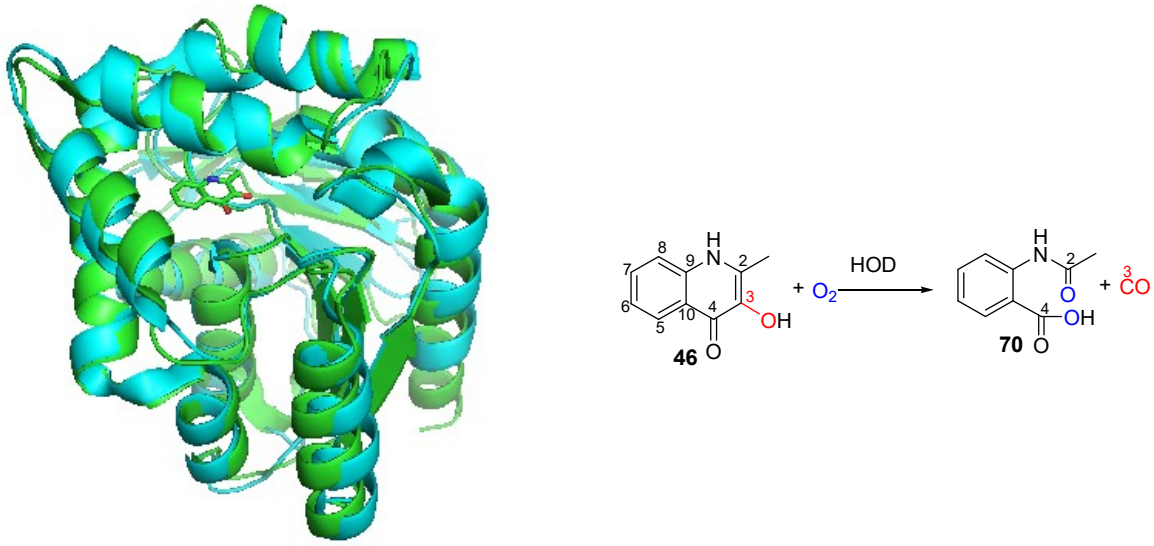
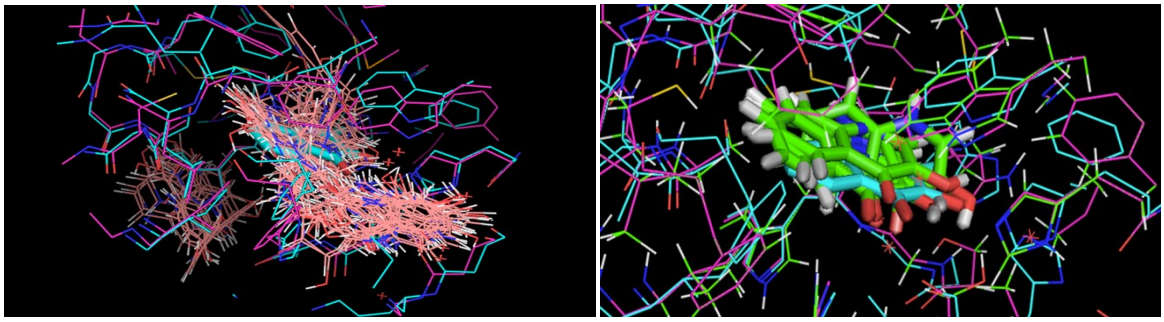


Figure 3.7: A) The overlap of HOD+**46** complex (green, PDB: 2WJ4) crystal and QDO crystal is shown (cyan, PDB: 3IBT); B) HOD catalyzes a similar reaction as QDO using substrate **46.**

3.2.3 MM optimization of the opened W153G QDO active site

The W153 residue of QDO was replaced by a glycine residue and then an MM calculation was performed on the mutated protein to find the optimized geometry with the lowest energy. The W153G QDO active-site (protein surface display) clearly showed a much larger opening than the wild type QDO active site (Figure 3.9). In addition, the Trp153 of QDO (W160 of HOD) might further stabilize the acylium ion via π-cation interaction (Figure 3.10, shown in HOD+**46** complex crystal structure, 2WJ4).

A) B)



C)

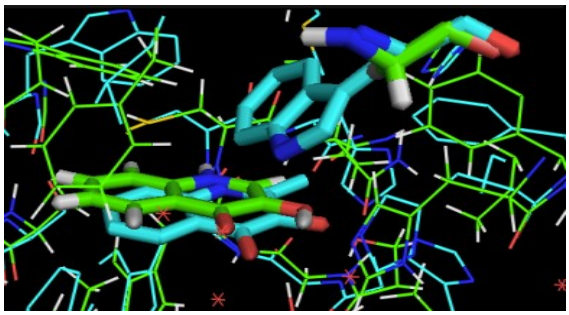


Figure 3.8: A) Flexible docking **1** into MM optimized QDO W153G (in magenta) using HOD+**46** structure (PDB: 2WJ4, in cyan) as the reference. 81 poses are shown in pink line; B) 5 (in green stick) out of the 81 poses are best overlaid with **46** in 2WJ4 (in cyan stick). The QDOwt structure (PDB: 3IBT) is also superimposed in green stick; C) The best conformation of docking **1** into the QDO binding pocket (in green stick) in comparison with **46** in HOD crystal structure (2WJ4, in cyan stick).

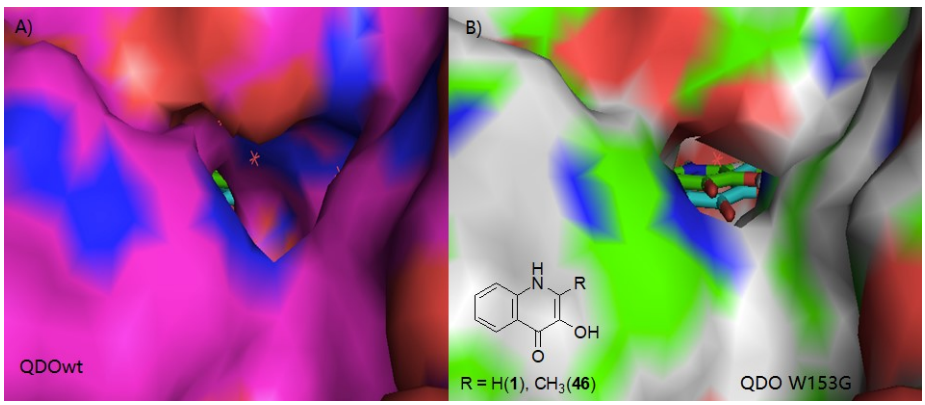


Figure 3.9: A) The surface view of the entrance of QDOwt binding pocket (PDB: 3IBT); B) The same viewing angle of the optimized QDO W153G binding pocket entrance. The QDO substrate **1** is in green, and the overlapped HOD substrate **46** is in cyan.

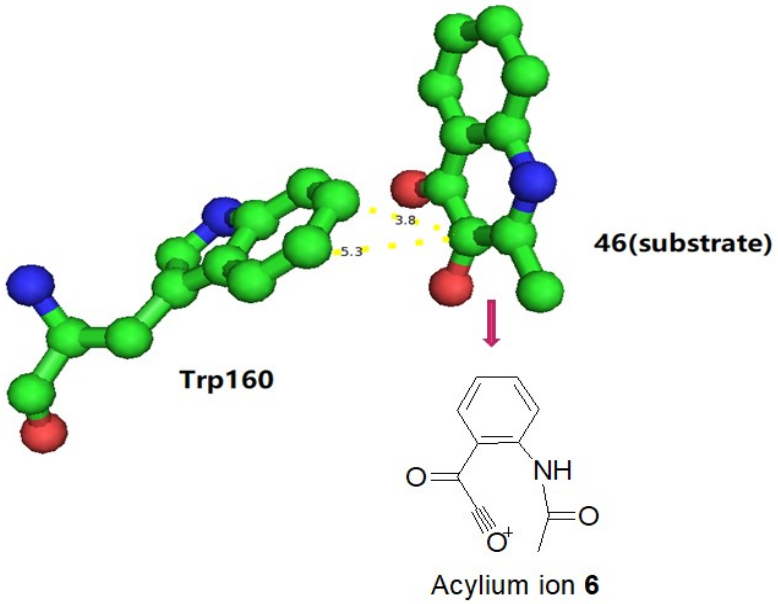


Figure 3.10: The orientation of Trp160 and the HOD substrate **46** in a crystal structure (2WJ4) might indicate the formation of π-cation interaction.

3.3 Conclusion

The DFT method was used to simulate energy profiles of the important intermediates and transition states of the alternative reaction mechanisms (formed either the cyclic peroxide intermediate or the acylium ion intermediates) of the QDO catalyzed reaction. Reasonable energy profiles were obtained for both mechanisms even without any protein residues around. The activation barrier generated through the acylium ion mechanism was 4 kcal/mol higher than through the cyclic peroxide mechanism, though they both had feasible activation energy barriers below 30 kcal/mol in solution. The barrier could be lower when the reaction happened inside the QDO binding pocket. The MM optimized geometry of the W153G QDO protein structure with the docking substrate **1** showed a larger opening of the binding pocket to the environmental

water, which facilitated the active site solvent exchange effect. All the calculation results matched with the experimental data very well.

3.4 Materials and methods

DFT Computational details: All structures were fully optimized in the gas-phase using Density Functional Theory (DFT) with the B3LYP (Becke-3 hybrid exchange[62] and Lee-Yang-Parr correlation) functional and a triple-ζ quality Pople-style basis set including diffuse and polarization functions (6-311+G(d,p))[63–65] on all atoms (BS1). Frequency calculations were performed to confirm the type of stationary point obtained (minimum or transition state). To account for solvation effects, Single Point Energy (SPE) implicit solvation calculations (water and dimethylformamide) were performed on the gas-phase ultrafine grids optimized geometry using the SMD parameterization[66] of the PCM[67] solvation model. All calculations were performed using the Gaussian 09 (G09) suite of programs[68]. The Cartesian coordinates for optimized DFT structures of the QDO calculations were shown in Appendix.

QDO substrate docking into the binding pocket and the W153G structure prediction: QDO substrate docking and W153G site-directed mutant geometry MM optimization were done using Discovery studio v4.1. The crystal structure of QDO (PDB ID: 3IBT, with a resolution of 2.6 Å) was used as the initial geometry. The protein was prepared by removing the water and some other co-crystallized small molecules, and potentials were assigned using CHARMm force field, the missing atom residues were built using the 'Build and Edit Protein' module, and the cleaning protein was prepared in 'Prepare Protein module'. The protein structure was energy minimized for 1000 steps (with the heavy atoms constrained) using the conjugate gradient algorithm with the 'Minimize and Refine Protein' module. The active site in

QDO was defined by overlapping QDO (PDB: 3IBT) and the active site of HOD+**46** (PDB: 2WJ4). The corresponding important active site residues of QDO were H244, H96, S95, W153, I185, Q32, and W30. During the docking process, the top 100 conformational poses were saved based on the ligand docking score value after the energy minimization through the LibDock module.

4. REAGENTS FOR THE DETECTION OF PROTEIN THIOCARBOXYLATES IN THE BACTERIAL PROTEOME AND THE OPTIMIZATION OF THE LABELING CONDITIONS

4.1 Protein thiocarboxylation PTM and proteomic approaches

Post-translational modification (PTM) is the covalent protein modification (usually enzymatic) after ribosomal mRNA-polypeptide translation step. PTMs can take place at the protein side chains, N- or C- terminals by nucleophilic addition, redox reaction, radical chemistry, etc. PTMs are widespread in both prokaryotes and eukaryotes and dramatically increase the diversity of the protein functions. Common examples of PTM include phosphorylation, glycosylation, methylation, acetylation, lipidation, and ubiquitination. Their functions include signaling, structure support, regulation, and so on. Thiocarboxylation is an unusual PTM that was only found at the C-terminal of the activated ThiS-like sulfur carrier proteins. ThiS-like sulfur carrier proteins are widely distributed in prokaryotes, and recently they are also reported in eukaryotes.[69] They are small proteins (~10kDa) that have structural similarity with ubiquitins. They have a flexible C-terminal tail ending with the featured Gly-Gly residues. The terminal glycine carboxylate group is first activated by an ATP-dependent enzyme to form an AMP phosphoester good leaving group. Then, the sulfur donors RSS^- or HS^- will attack the carboxylate group and form a protein thiocarboxylate (Figure 4.1).[70,71] ThiS-like proteins are efficient sulfur donors that play an important role in sulfur trafficking in the living organisms. Their sulfur acceptors can be very different, including RNA nucleotides, amino acids, sugars, cofactors and other sulfur-containing natural products (Figure 4.2).

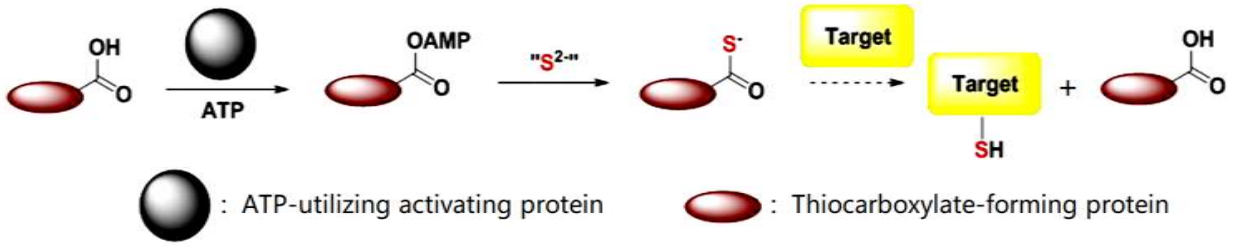


Figure 4.1: The biosynthesis of protein thiocarboxylates

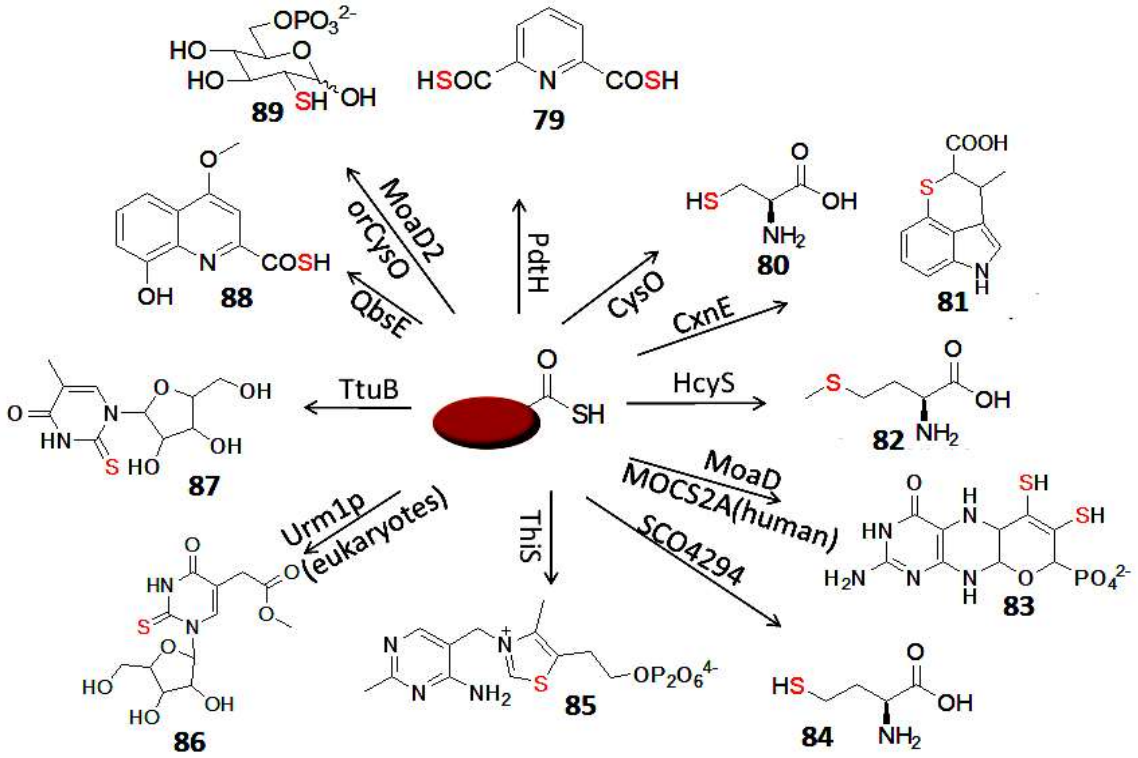


Figure 4.2: Summary of the reported ThiS-like sulfur carrier proteins involved in various natural product biosynthesis pathways. **79**: Pyridine dithiocarboxylic acid[21]; **80**: Cysteine[72]; **81**:Chuangxinmycin[73]; **82**: Methionine[26]; **83**: Molybopterin[74]; **84**: Homocysteine; **85**: Thiamin[75]; **86**: 5-methoxy-carbonyl-methyl-2-thiouridine[69]; **87**: 2-thioribothymidine[76]; **88**: Thioquinolobacterin[77]; **89**: 2-sulfur-glucose-6-phosphate[78].

Sequence analysis suggests that the ThiS-like proteins are widely distributed in bacteria and may participate in additional biosynthetic pathways.[21] Therefore, reagents that can

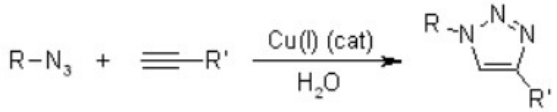
selectively interact with the protein thiocarboxylates in any proteome sample may be used to develop robust tools in direct visualizing or enriching the active ThiS-like protein thiocarboxylates from the proteome for further characterization of their biochemical functions. Moreover, even without bioinformatic context, these reagents may also be able to work independently in discovering the proteins with novel thiocarboxylation PTM sites other than the ThiS-like proteins. Also, these reagents may be used to probe the small thiocarboxylate metabolites for the metabolomics study.

Thiocarboxylates are strong nucleophiles in the physical environment. The selectivity of the labeling reagents may still be the key challenge for the biological thiocarboxylates detection when the thiocarboxylate functional group is usually in low concentrations with many other nucleophiles surrounded. We have focused our exploration on the "click chemistry" between the thiocarboxylates and the electron-deficient sulfonyl azides. [79–81]

4.2 Sulfonyl azide-thiocarboxylate click chemistry and the possible non-specific labeling in the proteome

"Click chemistry" is a concept introduced by Barry Sharpless in 2001, that refers to chemical reactions that react quickly (done by a "click") in mild conditions (e.g. neutral aqueous solution, better if biocompatible) with high yield and stereospecificity (i.e. with little or no side reactions) and requires simple or no work up steps.[82] A typical example of "click reaction" is the azide-alkyne cycloaddition that can ocurr rapidly with mild heating or Cu (I) catalyzing in aqueous solution.[83] Bertozzi's group has developed the copper-free click chemistry using the strained alkyne to react with the azide compound, which allows the 'click' reaction to be used in vivo with low toxicity (Figure 4.3).[84]

A)



B)

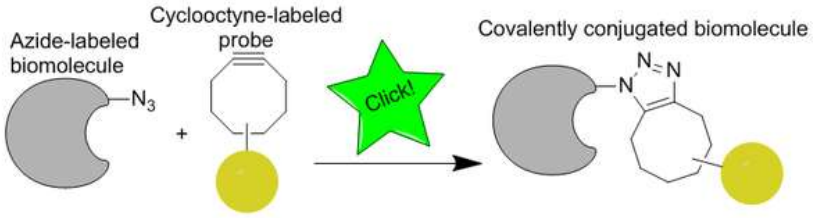


Figure 4.3: A) The copper(I) dependent and B) the copper-free azide-alkyne cycloaddition "click chemistry".

Williams' group has developed a new type of click chemistry which uses an electron-withdrawing azide to rapidly react with a thiocarboxylate compound at room temperature in aqueous solution without any other catalyst addition.[79–81] It couples the starting azide and thiocarboxylate together by forming a stable amide through the proposed mechanism in Figure 4.4. This reaction has been used to label the purified ubiquitin thiocarboxylate, indicating its potential for the proteomics investigation.[85]

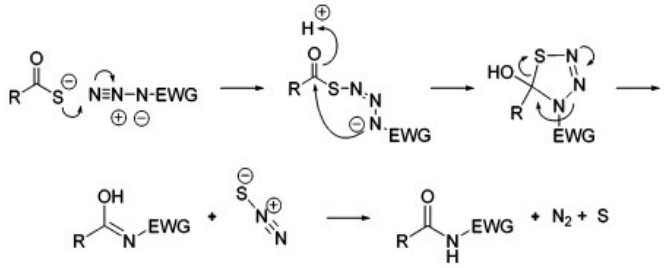


Figure 4.4: Proposed reaction mechanism of the "click chemistry" between a thiocarboxylate and an electron-deficient azide

We finally explored the aryl sulfonyl azids as the biological thiocarboxylate coupling reagents (Figure 4.5). These electron-deficient azides were designed to link with a rhodamin (**90**) or dansyl (**92**) group or a biotin (**91**) group to selectively report the thiocarboxylates by

fluorescence imaging or enrich the low concentration of thiocarboxylates in the bacterial cell extract for further characterization using LC-MS. Two previous Ph.D. students from our lab, Kalyanaraman Krishnamoorthy and Yiquan Liu, have already used the rhodamine and biotin linked aryl sulfonyl azide reagents (**90** and **91**) to selectively label the two ThiS-like sulfur carrier proteins involved in the biosynthesis of pyridine dithiocarboxylic acid (**79**)[21] and homocysteine (**84**) separately in the bacterial proteomes (Figure 4.2).

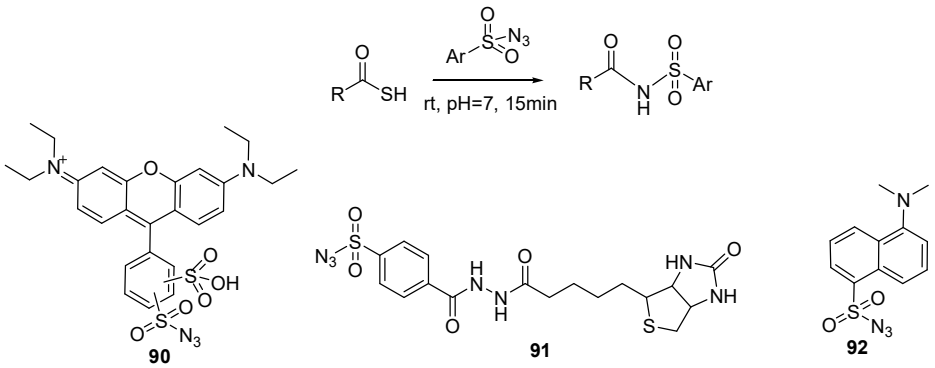


Figure 4.5: The "click reaction" between aryl sulfonyl azides (**90-92**) and the thiocarboxylates (both protein thiocarboxylates and small thiocarboxylate metabolites). **90**: Lissamine Rhodamine B Sulfonyl Azide (LRSA); **91**: biotin-4-carboxybenzene sulfonyl azide (BiotinSA); **92**: dansyl azide. LRSA (**90**) are mixtures of isomers (either 2-sulfonyl acid-4-sulfonyl azide, or 2-sulfonyl azide-4-sulfonyl acid)

Besides thiocarboxylates, other nucleophiles that broadly spread in the proteome, such as cysteine and lysine, may also be able to interact with the sulfonyl azides (Figure 4.6B-C). Kalyan has used TCEP to remove the cysteine-LRSA non-specific labeling.[21] In this chapter, I describe the condition optimization on minimizing the lysine-LRSA non-specific binding.

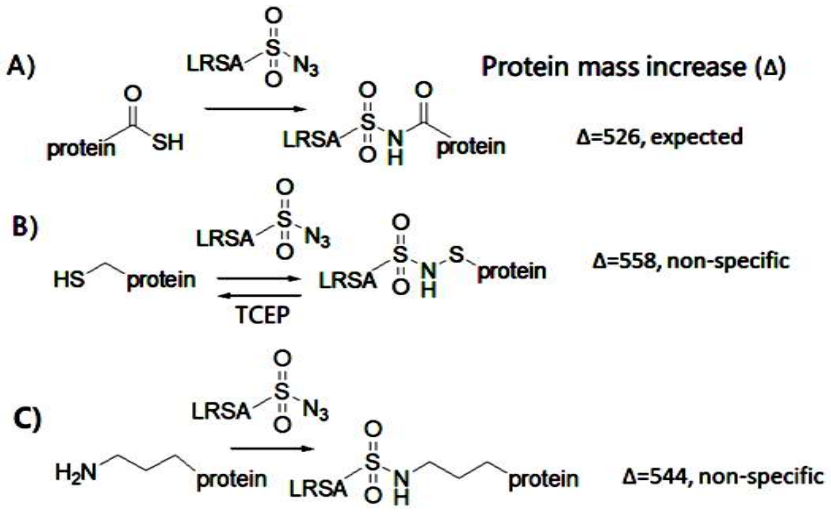


Figure 4.6: Common nucleophiles found in a proteome that can interact with LRSA (**90**). A) LRSA (**90**) reacts with protein thiocarboxylates (the target reaction). The protein mass will increase 526 Da after LRSA labeling. B) LRSA (**90**) reacts non-specifically with cysteine residues. The cysteine labeled protein mass will increase 558 Da. This adduct can be converted back to cysteine through TCEP reduction. C) LRSA (**90**) reacts non-specifically with the lysine residue. The lysine labeled protein mass will increase 544 Da.

4.3 Results and discussions

4.3.1 Model reaction: LRSA (90) interaction with a small thiocarboxylate (thioacetic acid, 93) and high concentration of a simple primary amine (2-methoxyethanamine, 97)

To figure out the chemistry details, I simplified the protein thiocarboxylation site to a small thiocarboxylate (thioacetic acid, **93**), and the lysine residue to a simple primary amine (2-methoxyethanamine, **97**) (Figure 4.7). The LRSA (**90**) was first interacted with the thioacetic acid (**93**) alone and quenched with TCEP (Figure 4.8). Since LRSA (**90**) are isomer mixtures of 2-sulfonylazide-4-sulfonylacid and 2-sulfonylacid-4-sulfonylazide, the LC-MS data showed five product peaks corresponding to the desired LRSA-thiocarboxylate adduct isomers (**94**, [M+1] = 600), the LRSA hydrolyzed product (**95**, [M+1] = 559), and the TCEP reduction product isomers (**96**, [M+1] = 558).

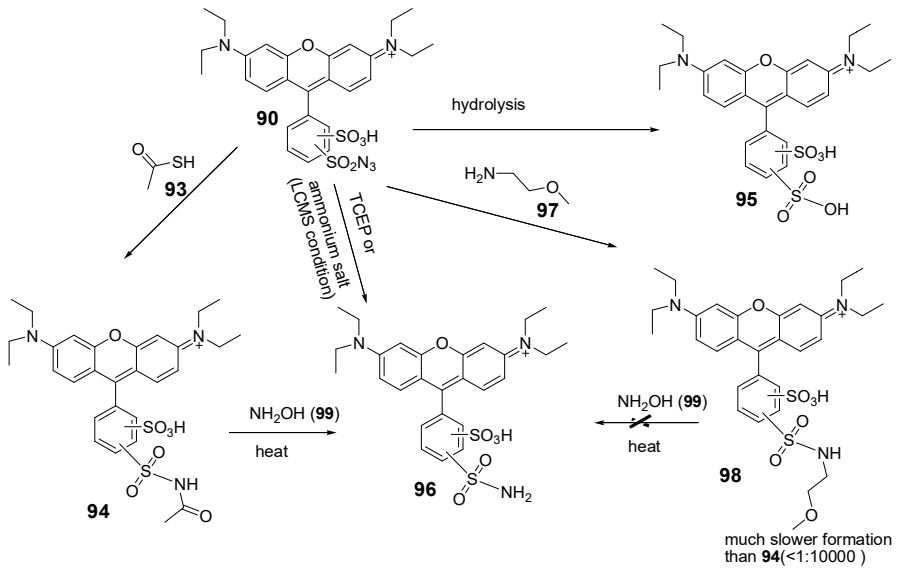


Figure 4.7: The scheme of LRSA (**90**) interaction with thioacetic acid (**93**) and a simple amine (2-methoxyethanamine, **97**) and other side reactions under the protein thiocarboxylate labeling conditions.

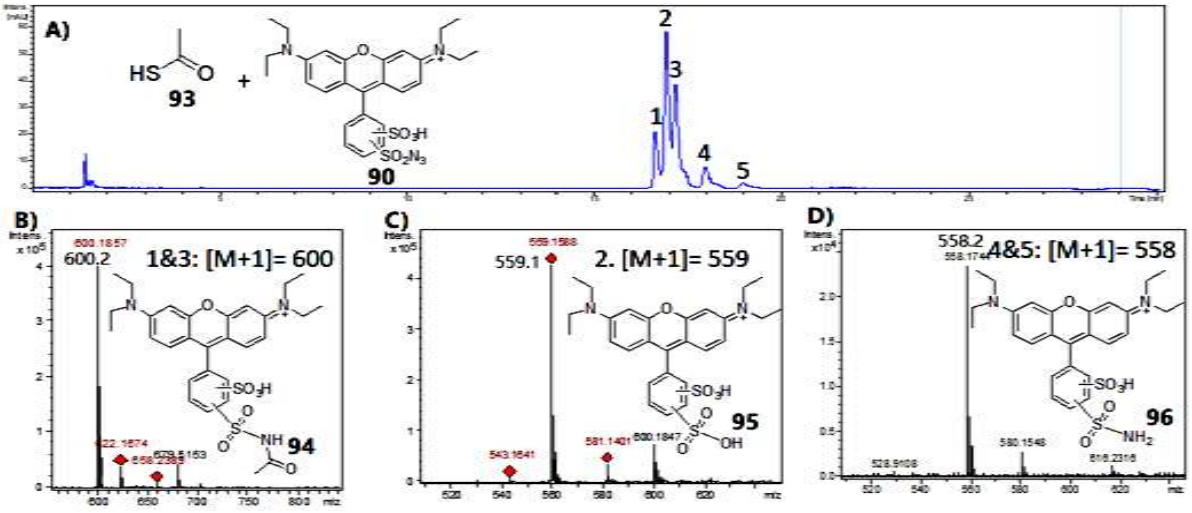


Figure 4.8: LC-MS analysis of LRSA (**90**) interaction with thioacetic acid (**93**) for 15 min. TCEP was used to quench the reaction by reducing the excess amount **90** to **96**. A) LC of the reaction mixture at 567nm (the specific absorption wavelength of LRSA (**90**)). The product peaks 1-5 were characterized by high-resolution ESI-MS analysis in B)-D). Since LRSA (**90**) are isomer mixtures of 2-sulfonylazide-4-sulfonylacid and 2-sulfonylacid-4-sulfonylazide, Peaks 1 and 3 had the same mass ([M+1] =600, see B)), corresponding to the 2- and 4- isomers of the LRSA-thioacetic acid adduct (**94**). Peak 2 was the hydrolyzed product (**95**) of LRSA (**90**) ([M+1]=559, see C)). Peak 4 and 5 were the LRSA (**90**) reduced product isomers (**96**) ([M+1] = 558, see D)) by TCEP.

Then the thioacetic acid (**93**) was labeling with LRSA (**90**) with the existence of 2-methoxyethanamine (**97**). Considering the abundance of lysine side chain amines in the proteome is much higher than the protein thiocarboxylates, a mixture contained 1:100 ratio of thioacetic acid (**93**) and 2-methoxyethanamine (**97**) in varied pH buffer was used for the LRSA (**90**) labeling (Figure 4.9). The pKa of the lysine side chain is 10.5, so the lysine amine is protonated and becomes a poor nucleophile at neutral to acidic pH. In contrast, the thiocarboxylate group is considered as fully deprotonated in neutral condition and therefore it shows much stronger nucleophilicity than lysine at neutral pH. Mixed with 100 fold higher concentration of 2-methoxyethanamine (**97**), LRSA (**90**) still interacted with thioacetic acid (**93**) rapidly in 15 min while no LRSA-amine adduct (**98**) was observed at 567 nm (the LRSA special absorbance wavelength). In fact, the LRSA-amine adduct isomers (**98**) were formed in small amount which could only be detected by the more sensitive mass searching method (the extracted ion chromatogram, EIC, see Figure 4.9C). In the mixture, the thiocarboxylate (**93**) and amine (**97**) initial concentration was 1: 100. The LC product peaks area comparison (at 567 nm) suggested that the rough yield ratio of LRSA-thiocarboxylate adduct (**94**): LRSA-amine adduct (**98**) was much higher than 100:1. Therefore, the LRSA (**90**) had more than 10000 times selectivity on labeling the thiocarboxylate **93** than labeling the amine **97**. In the proteome conditions, the interaction between amino acid residues may change the local environment of certain lysine amines and make it easier to lose a proton and become a better nucleophile. The concentration of lysine residues is also much higher than the protein thiocarboxylates. The effect of the non-specific LRSA-lysine interaction may be more significant to the LRSA-

thiocarboxylate interaction selectivity in the proteomics study. The LRSA (**90**) labeling efficiency was tested in various pH buffer (pH 5, 6 and 7). In pH 7 PBS buffer, LRSA (**90**) generated the most LRSA-thiocarboxylate adduct (**94**) and the least LRSA-amine adduct (**98**). The lower pH did not reduce the lysine non-specific labeling but decreased the LRSA-thiocarboxylate labeling efficiency. Thus, the following labeling tests were all done in pH 7 buffer.

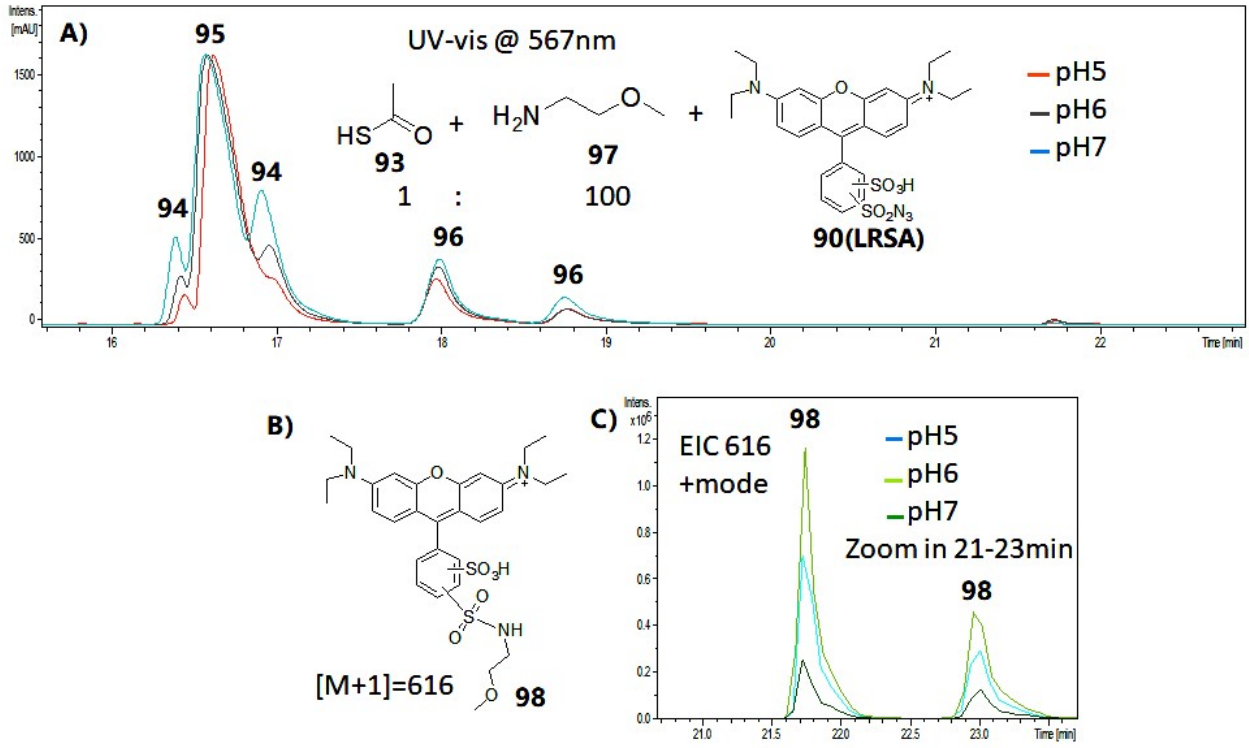


Figure 4.9: LC-MS analysis of the reaction mixture of LRSA (**90**), thioacetic acid (**93**) and 2-methoxyethanamine (**97**) after incubation for 15 min in various pH. The concentration ratio of **93** and **97** is 1:100. TCEP was used to quench the reaction. A) The LC traces (567 nm) of the reaction mixture in pH 5-7 buffer. The LRSA (**90**)- 2-methoxyethanamine (**97**) adduct (**98**) was not observed at 567 nm but could only be detected by EIC (extracted ion chromatogram, see C)). B) The structure of **98** with [M+1] = 616. C) The EIC = 616 for **98** isomers detection from the same data sets in A). The optimized pH was pH 7. The rough estimation of the product ratio of **94**:**98** was > 10000:1.

The LRSA-thiocarboxylate adduct (**94**) and the LRSA-amine adduct (**98**) had distinguishable properties when adding the hydroxylamine (NH_2OH, **99**) and heating at 100 °C (Figure 4.7 and 4.10). **94** was fully converted to **96** with the nucleophilic attack of NH_2OH under 15 min heating at 100 °C. Meanwhile, **98** did not interact with NH_2OH (**99**) after heating for 15 min at 100 °C.

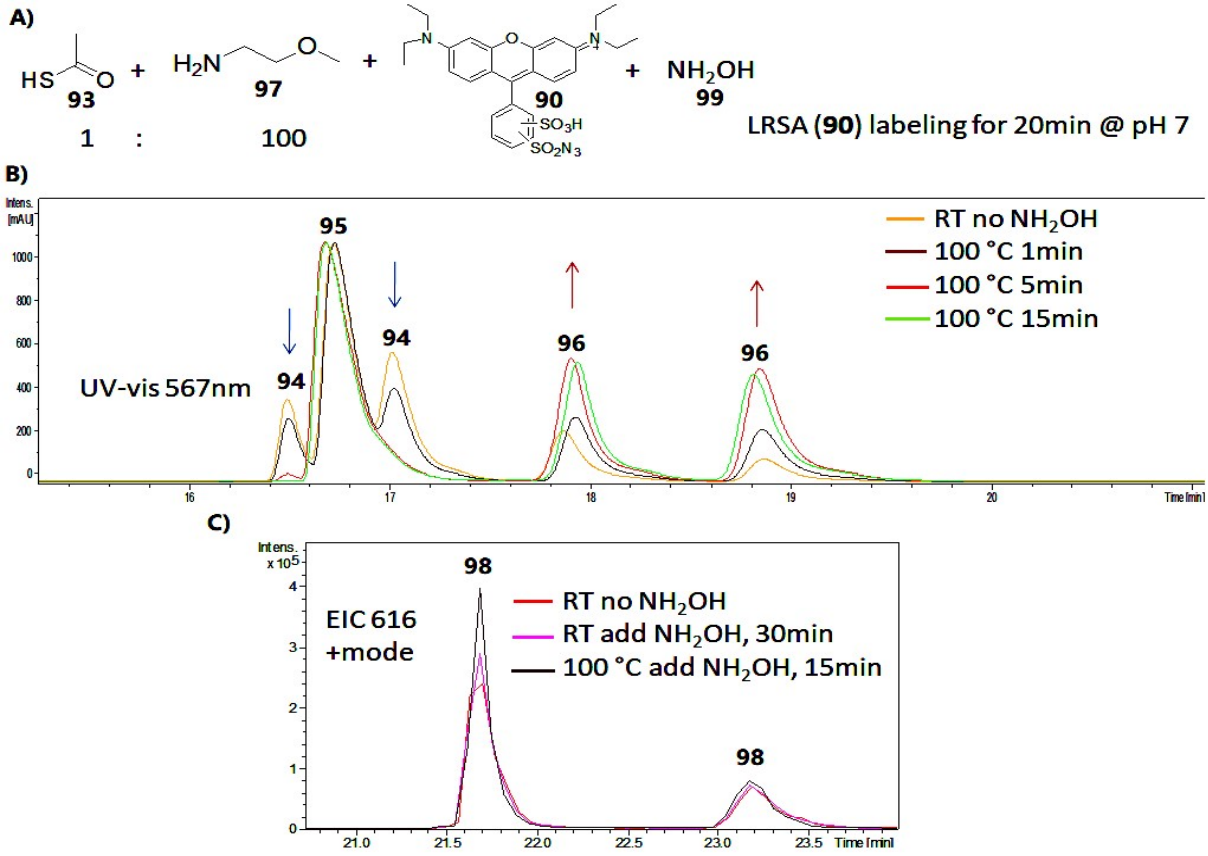


Figure 4.10: The LC-MS analysis of NH_2OH (**99**) treatment of the reaction mixture of LRSA (**90**), thioacetic acid (**93**) and 2-methoxyethanamine (**97**). A) The structure of reactants. A mixture of **90**, **93** and **97** were incubated in the dark for 20 min at RT and quenched by TCEP. NH_2OH (**99**) was then added to the reaction mixture in various conditions. B) The LC at 567 nm of the reaction mixture under different NH_2OH (**99**) treatment conditions. All the LRSA-thioacetic acid adduct (**94**) was transformed into **96** after the addition of NH_2OH (**99**) and heating for 15 min at 100 °C. C) EIC = 616 for the detection of the LRSA-amine adduct (**98**) after various conditions of NH_2OH (**99**) treatment. No **98** decomposition was observed after the addition of NH_2OH and heating for 15 min at 100 °C.

4.3.2 NH_2OH (99) treatment to distinguish the LRSA (90)-protein thiocarboxylate labeling from the non-specific lysine labeling in purified proteins and bacterial proteomes

NH_2OH was further used to determine the "fake positive" LRSA-thiocarboxylate labeling in bacterial proteome samples. As a positive control, ThiS-COSH from *Thermus thermophilus* was overexpressed and purified[21] for LRSA (**90**) labeling and NH_2OH treatment (Figure 4.11). After LRSA (**90**) labeling and TCEP quenching (Figure 4.11, lane 3), NH_2OH (**99**) was added to the sample and heated for 20 min at 100 °C (Figure 4.11, lane 4). The fluorescence intensity of ThiS-COSH in lane 4 was significantly decreased compared with lane 3 and the released free rhodamine band was also found in lane 4.

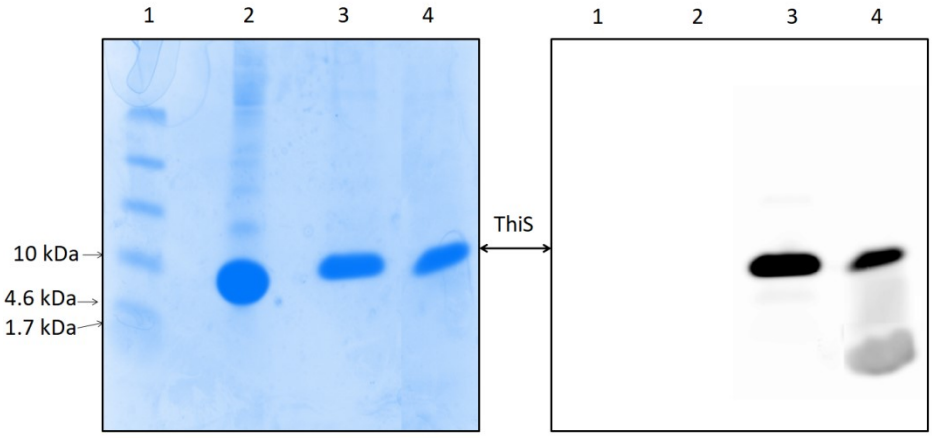


Figure 4.11: The Tris-Tricine gel analysis of the NH_2OH treatment after ThiS-COSH (*T. thermophilus*) labeled by LRSA (**90**). Left: the Coomassie blue stain; Right: the fluorescence image at Ex/Em 530/580 nm, the LRSA (**90**) fluorescence range. Lane 1: the protein ladder; lane 2: the purified ThiS-COSH; lane 3: the purified ThiS-COSH labeled by LRSA (**90**); lane 4: the NH_2OH (**99**) treatment (heat at 100 °C for 20 min) after ThiS-COSH labeled by LRSA (**90**). The band with molecular weight lower than 1.7 kDa was the free **96**.

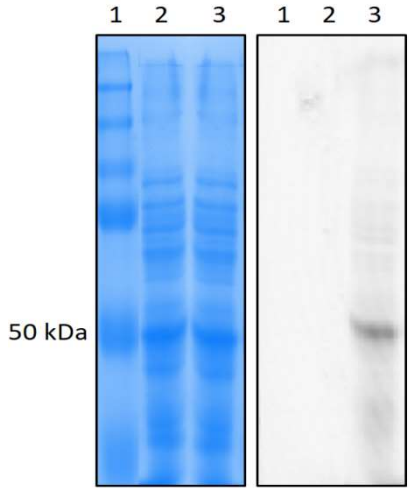


Figure 4.12: The SDS-PAGE analysis of the LRSA (**90**) labeled *E. coli BL21 (DE3)* proteome. Left: the Coomassie blue stain; Right: the fluorescence image at Ex/Em 530/580 nm. Lane 1: the protein ladder; lane2: the *E. coli BL21 (DE3)* proteome; lane 3: the *E. coli BL21 (DE3)* proteome labeled by LRSA (**90**) for 30 min in the dark at RT and quenched by TCEP. A protein band at 50 kDa seemed to be labeled by **90**.

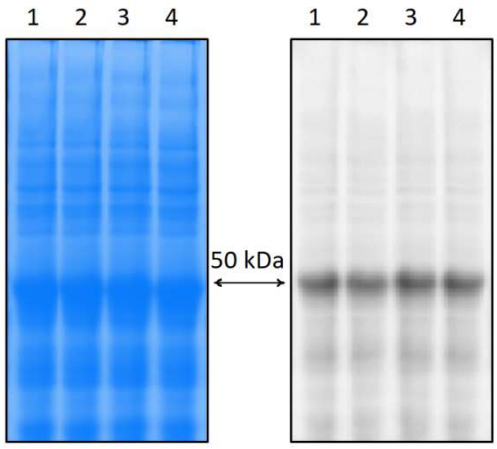


Figure 4.13: The SDS-PAGE analysis of the NH_2OH (**99**) treatment of the LRSA (**90**) labeled *E. coli BL21 (DE3)* proteome. Left: the Coomassie blue stain; Right: the fluorescence image at Ex/Em 530/580 nm. The proteome was incubated in excess of LRSA (**90**) in the dark for 30 min, quenched with TCEP, and then added NH_2OH (**99**) and heated at 100 °C for various time periods. Lane 1: E. coli proteome labeled by LRSA (**90**) for 30 min, quenched by TCEP. The same sample of lane 1 treated with NH_2OH (**99**) and heated at 100 °C for 15 min (lane 2), 30 min (lane 3) and 45 min (lane 4). The 50 kDa band labeling maintained even after 45 min heating at 100 °C.

When I repeated Kalyan's procedure[21] to label the *E. coli* proteome, I found a fluorescence band at 50 kDa was much darker than the non-specific labeling background (Figure 4.12). The Coomassie blue staining showed that the corresponding protein band was in much higher concentration than other protein bands. To figure out whether the band was a new protein thiocarboxylate or was just due to the accumulation of lysine non-specific labeling, the LRSA (**90**) labeled E. coli proteome was treated with NH_2OH and heated for 15 min, 30 min, and 45 min at 100 °C (Figure 4.13). The fluorescence intensity of the labeled band at 50 kDa maintained the same even after the addition of NH_2OH and heating for 45 min at 100 °C. This result suggested that the labeled band at 50 kDa in E. coli proteome might be a "false positive" caused by the LRSA-lysine non-specific labeling. Besides the protein band at 50 kDa, other much lighter non-specific labeling bands were not faded either, which indicated that maybe most non-specific labeling background was caused by the LRSA-lysine interaction.

4.3.3 Optimizing the LRSA (90) labeling conditions for efficient protein thiocarboxylate labeling and minimized non-specific lysine labeling in the bacterial proteome

The previous small molecule model labeling studies showed that the LRSA-lysine interaction is much slower than the LRSA-thiocarboxylate reaction. To minimum the non-specific lysine labeling while still maintaining the labeling efficiency of protein thiocarboxylates, one of the most straightforward way was to find the proper LRSA (**90**) labeling incubation time. LRSA (**90**) was incubated with the E. coli proteome in the dark for 10 min, 20 min, 30 min and 1 hr at RT before the TCEP quenching and desalting steps (Figure 4.14). When the labeling time was 10 min, there was barely any non-specific background labeling. When the labeling time increased, the non-specific labeling background also increased accordingly. The darker band at

50 kDa was cut from the SDS-PAGE gel and sent for in-gel peptide digestion nanoLC-MS analysis and the most abundant protein in this gel-cut band was tryptophanase (TnaA) (Figure 4.15).

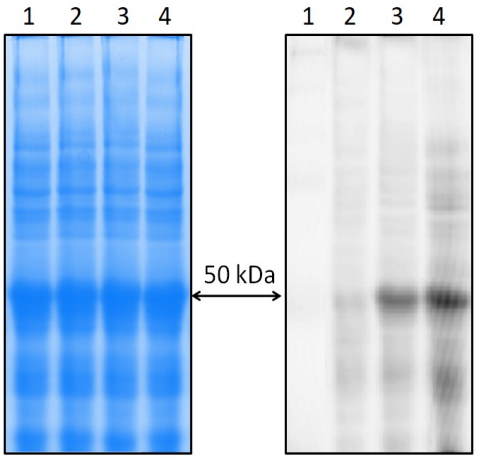


Figure 4.14: The SDS-PAGE analysis of the *E. coli* proteome incubated with LRSA (**90**) for various time periods before the TCEP quenching. Left: the Coomassie blue stain; Right: the fluorescence image at Ex/Em 530/580 nm. Lane 1: the *E. coli* proteome incubated with LRSA (**90**) for 10 min; lane 2: for 20 min; lane 3: for 30 min; lane 4: for 1 hr.

This enzyme is a PLP-dependent enzyme which has an active-site lysine to interact with PLP through nucleophilic addition (Figure 4.16). Before the labeling process, the proteome was solvent-exchanged into 6 M urea buffer and stayed for 15 min to unfold the proteins. Some proteins may partially maintain their secondary structure. The lysine might be more nucleophilic to attack LRSA (**90**) than free amines. The tryptophanase (TnaA) was cloned, overexpressed and purified for LRSA (**90**) labeling (30 min) and NH_2OH (**99**) treatment and the SDS-PAGE analysis showed the same labeling pattern as the E. coli proteome labeling at 50 kDa (Figure 4.17 and Figure 4.13). The fluorescence intensity did not decrease after NH_2OH/heat treatment due to the LRSA-lysine adduct formation.

1	Description	Coverage	# Peptides	# PSMs	# Unique Peptides	# Protein Groups	# AAs	MW [kDa]	calc. pI	Score Sequest HT	# Pept
2	Tryptophanase OS=Escherichia coli O157:H7 GN=tnaA PE=3 SV=1	34.60721868	15	77	15	1	471	52.756	6.23	205.9999156	
3	Tryptophanase OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=tnaA PE=:	34.60721868	15	77	15	1	471	52.768	6.23	205.9999156	
4	Tryptophanase OS=Escherichia coli O9:H4 (strain HS) GN=tnaA PE=3 SV=1	34.60721868	15	77	15	1	471	52.798	6.07	205.9999156	
5	Tryptophanase OS=Shigella boydii serotype 18 (strain CDC 3083-94 / BS512) GN=tnaA PE=3 SV	34.60721868	15	77	15	1	471	52.809	6.39	205.9999156	
6	Tryptophanase OS=Shigella flexneri serotype 5b (strain 8401) GN=tnaA PE=3 SV=1	34.60721868	15	77	15	1	471	52.74	6.23	205.9999156	
7	Tryptophanase OS=Escherichia coli (strain SMS-3-5 / SECEC) GN=tnaA PE=3 SV=1	31.63481953	14	74	14	1	471	52.708	6.23	195.8854856	
8	Tryptophanase OS=Escherichia fergusonii (strain ATCC 35469 / DSM 13698 / CDC 0568-73) GN=	29.08704883	13	73	13	1	471	52.674	6.23	192.5036135	
9	Glycerol kinase OS=Escherichia coli (strain 55989 / EAEC) GN=glpK PE=3 SV=1	41.63346614	18	25	6	1	502	56.195	5.5	79.25644827	
10	Glycerol kinase OS=Shigella boydii serotype 18 (strain CDC 3083-94 / BS512) GN=glpK PE=3 SV	41.63346614	18	25	6	1	502	56.136	5.8	79.25644827	
11	Glycerol kinase OS=Escherichia coli O8 (strain IAI1) GN=glpK PE=3 SV=1	40.03984064	17	24	5	1	502	56.209	5.5	77.03412485	
12	Glycerol kinase OS=Escherichia coli O7:K1 (strain IAI39 / ExPEC) GN=glpK PE=3 SV=1	38.24701195	17	24	6	1	502	56.179	5.5	74.14916801	
13	Glycerol kinase OS=Shigella flexneri GN=glpK PE=3 SV=2	39.64143426	17	23	6	1	502	56.176	5.5	73.98310995	
14	Glycerol kinase OS=Escherichia coli O81 (strain ED1a) GN=glpK PE=3 SV=1	37.05179283	16	23	4	1	502	56.268	5.5	73.16946459	
15	Glycerol kinase OS=Escherichia coli O127:H6 (strain E2348/69 / EPEC) GN=glpK PE=3 SV=1	37.05179283	16	23	4	1	502	56.254	5.5	73.16946459	
16	Glycerol kinase OS=Escherichia fergusonii (strain ATCC 35469 / DSM 13698 / CDC 0568-73) GN	37.05179283	16	23	4	1	502	56.226	5.5	73.16946459	
17	Glycerol kinase OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPEC) GN=glpK PE=:	37.05179283	16	23	4	1	502	56.24	5.5	73.16946459	
18	Glycerol kinase OS=Shigella dysenteriae serotype 1 (strain Sd197) GN=glpK PE=3 SV=1	38.44621514	17	23	5	1	502	56.194	5.59	71.53864098	
19	Glutamate decarboxylase beta OS=Escherichia coli (strain K12) GN=gadB PE=1 SV=1	30.25751073	11	19	11	1	466	52.634	5.49	69.92932415	
20	Glutamate decarboxylase alpha OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPI	30.25751073	11	19	11	1	466	52.651	5.41	69.92932415	
21	Glutamate decarboxylase alpha OS=Shigella flexneri GN=gadA PE=3 SV=2	30.25751073	11	19	11	1	466	52.666	5.41	69.92932415	
22	Glutamate decarboxylase alpha OS=Escherichia coli O157:H7 GN=gadA PE=3 SV=1	26.60944206	9	16	9	1	466	52.665	5.41	56.90932393	
23	Glutamate decarboxylase beta OS=Escherichia coli O6:H1 (strain CFT073 / ATCC 700928 / UPE	26.60944206	10	16	10	1	466	52.619	5.58	56.68111134	
24	Glycerol kinase OS=Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696) GN=g	29.28286853	13	18	1	1	502	56.103	5.6	56.6297214	
25	Glycerol kinase OS=Salmonella agona (strain SL483) GN=glpK PE=3 SV=1	29.28286853	13	18	1	1	502	56.002	5.6	55.97031283	

Figure 4.15: The protein database searching of nanoLC-MS in-gel digested peptide analysis of the LRSA (**90**) labeled 50 kDa gel band cut. The highest scored protein (usually the most abundant) was tryptophanase (TnaA, red-lined), which is a PLP-dependent enzyme with an active-site lysine binding to the PLP aldehyde group through nucleophilic addition.

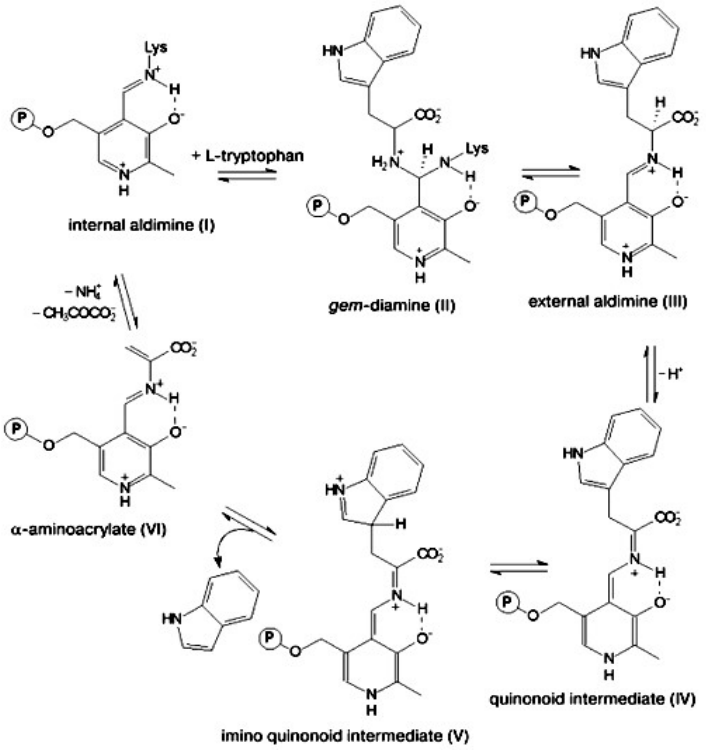


Figure 4.16: The proposed PLP-dependent mechanism of tryptophanase (TnaA).[86] The protein active-site lysine interacted with PLP aldehyde group to form an imine through nucleophilic addition.

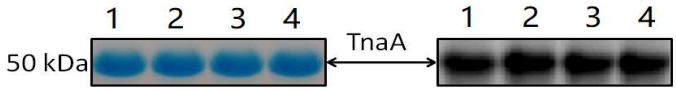


Figure 4.17: The SDS-PAGE analysis of the purified tryptophanase (TnaA) labeled with LRSA (**90**) for 30 min, quenched with TCEP, treated with NH_2OH (**99**) and heated at 100 °C in various time periods. Lane 1: LRSA (**90**) labeled TnaA without NH_2OH addition; lane 2: LRSA labeled TnaA with NH_2OH treatment and heated for 15 min; lane 3: labeled TnaA+NH_2OH heated for 30 min; lane 4: labeled TnaA+NH_2OH heated for 45 min

We also tested the protein thiocarboxylate efficient labeling time using purified ThiS-$COSH_{Tt}$ (Figure 4.18). LRSA (**90**) was incubated with ThiS-$COSH_{Tt}$ for 10 min, 30 min and 1hr, quenched with TCEP, desalted and loaded onto the SDS-PAGE gel. The fluorescence intensity was similar between 10 min and 30 min labeling time and a little higher when labeling for 1 hr

(maybe due to the non-specific lysine labeling). This result showed that LRSA (**90**) labeling ThiS-$COSH_{Tt}$ was very efficient and could be completed within 10 min.

Based on all the above controls, the LRSA (**90**) labeling incubation time was optimized to 15 min for efficiently labeling the protein thiocarboxylate while minimizing the non-specific lysine labeling. The proteome of wild-type *Lc. lactis NZ3900* (more labeling experiments using this strain described in Chapter 5) was labeled with LRSA (**90**) under this condition, and almost no non-specific labeling background could be found (Figure 4.19).

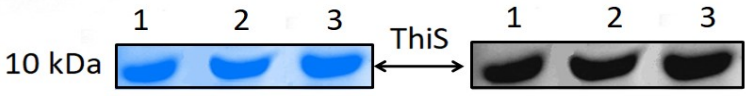


Figure 4.18: The SDS-PAGE analysis of the purified ThiS-COSH incubated with LRSA (**90**) for 10 min (lane 1), 30 min (lane 2) and 1 hr (lane 3) before TCEP quenching. Left: the Coomassie blue stain; Right: the fluorescence image at Ex/Em 530/580 nm.

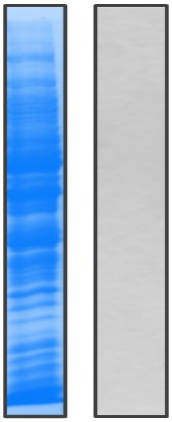


Figure 4.19: The SDS-PAGE analysis of LRSA (**90**) labeling the *Lc. lactis NZ3900* proteome for 15 min in the dark at RT (quenched by TCEP). Left: the Coomassie blue stain; Right: the fluorescence image at Ex/Em 530/580 nm. No significant non-specific labeling band was found in the fluorescence image.

4.4 Conclusion

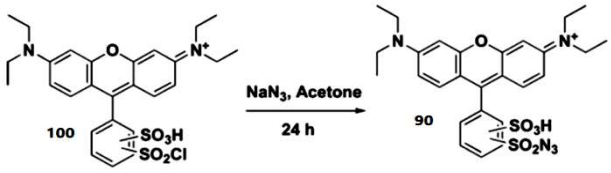
ThiS-like protein thiocarboxylates could be efficiently labeled with LRSA (**90**) in 10 min. The nucleophilic lysines in the proteome also interacted with LRSA (**90**) at the much slower rate. This non-specific lysine labeling was investigated using the small molecule model reaction, the purified proteins, and the bacterial proteomes. The LRSA (**90**) labeling time was optimized to assure the protein thiocarboxylate getting labeled completely as well as minimizing the non-specific lysine labeling. The *Lc. lactis* proteome was labeled with LRSA (**90**) for 15 min and showed a clean background with no visible non-specific fluorescent protein bands. With the improved selectivity, LRSA (**90**) might become a biorthogonal and sensitive reagent to fish new protein thiocarboxylates from the environmental proteome pools.

4.5 Materials and methods

Materials: All chemicals were purchased from Sigma-Aldrich unless otherwise mentioned. LB-Lennox broth was from EMD Millipore. Kanamycin and IPTG were obtained from Lab Scientific Inc. Chloramphenicol was from Fisher Scientific. D-chloroform, D_6-DMSO, D_4-methanol were purchased from Cambridge Isotope Laboratories. 2.5L baffled ultra yield flask for overexpressing proteins were purchased from Thomson Instrument Company. The fluorescence images were scanned with a Typhoon FLA 9500 instrument from GE Healthcare Biosciences (Piscataway, NJ). The OD_{600} and other UV-vis absorptions were measured using a Varian Cary 300 Bio UV-Visible Spectrophotometer (Palo Alto, CA). Tryptophanase (TnaA) cloned in pTHT vector was obtained from Dr. Xiaohong Jian, the manager of our protein facility. The *Lc. lactis NZ3900* strain used was kindly provided by Dr. Robert P. Hausinger's lab (Table 5.1).

Synthesis of Lissamine Rhodamine B Sulfonyl Azide (LRSA, 90)[21]: Lissamine rhodamine B sulfonyl chloride (53 mg, 92 µmol; compound **100**, Figure 4.20A) was dissolved in 10 mL of acetone in a round-bottomed flask wrapped in aluminum foil. Sodium azide (29 mg, 446 µmol, 5 equiv.) was added, and the reaction mixture was stirred at RT for 24 h. The solvent was removed in vacuo, and the residue was redissolved in dichloromethane. The resulting solution was washed with water, dried over anhydrous $MgSO_4$, and filtered, and the solvent was removed in vacuo to give LRSA (**90**; 46 mg, 85%). A 15 mM stock solution in dimethylsulfoxide (DMSO) was prepared and stored in 1 ml aliquots at -20 °C in the dark.

A)



B)

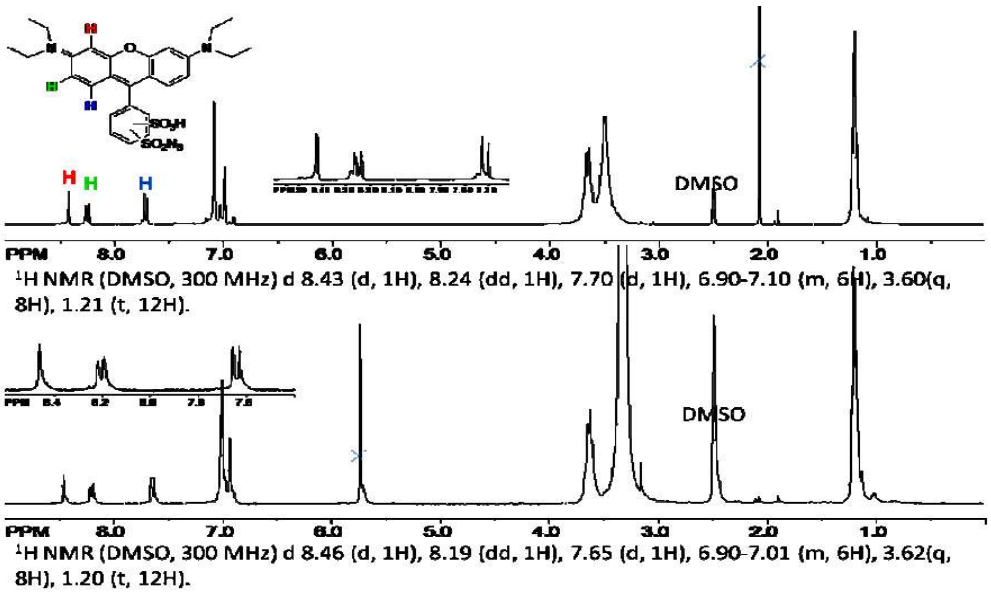


Figure 4.20: A) Synthesis of LRSA (**90**); B) ^{1}H-NMR spectra of the LRSA (**90**) regio-isomers – isomer A (top) and isomer B (bottom).

LRSA (90) interaction with thioacetic acid (93) and 2-methoxyethanamine (97), and the treatment of NH_2OH (99): LRSA (**90**) interaction with thioacetic acid (**93**): A 100 μL mixture of LRSA (**90**) and thioacetic acid (**93**) (final concentration: 2 mM **90** and 200 μM **93**) in the pH 7 PBS buffer was incubated in the dark at RT for 15 min and quenched by TCEP (final concentration: 10 mM). The mixture was placed in the dark for another 15 min. 50 μL of the mixture was taken for HPLC analysis and 25 μL of the mixture was used for LC-MS analysis.

LRSA (**90**) interaction with thioacetic acid (**93**) and 2-methoxyethanamine (**97**): A 100 μL mixture of LRSA (**90**), thioacetic acid (**93**) and 2-methoxyethanamine (**97**) (final concentration: 2 mM **90**, 200 μM **93** and 20 mM **97**) in the pH 5-7 PBS buffer was performed the same procedure as mentioned above and analyzed by HPLC and LC-MS.

The NH_2OH (**99**) treatment: The above reaction mixture was added 50 mM NH_2OH (**99**) and either incubated at the RT for 30 min, or immediately heated at 100 °C for 1 min, 5 min, 15 min. Then the samples were sent for HPLC and LC-MS analysis.

Overexpression and purification of E. coli tryptophanase (TnaA): TnaA was cloned in pTHT vector (derivative of the pET28b vector with TEV protease cleavage site after the N-terminal His-tag) overexpressed in *E.coli BL21(DE3)*. 1.5 L culture was grown at 37 °C in LB media until an OD_{600} of 0.6. The temperature was then reduced to 15 ^{o}C, and the culture was induced with 1 mM IPTG and continued growing at 15 ^{o}C for 12-16 hr with constant shaking (180 rpm). To purify the protein, the cell pellets from 1-1.5 L culture were mixed with 20-30 ml lysis buffer (50 mM KPi, 300 mM NaCl, pH 7) and thawed on ice with addition of 6-8 mg lysozyme for 1 hr. The cell suspension was stirred to homogeneous and sonicated (the Misonix Sonicator 3000 model, 1.5s on/1.5s off pulses for 30s intervals, 5-6 times, allowing 5 min

cooling time on ice between each interval) to lyse the cells thoroughly. The cell debris was removed by centrifugation at 15000 rpm for 30 min. The supernatant cell lysate was filtered through 0.45 μm filters and loaded onto a Ni-NTA His-trap column (equilibrated with lysis buffer). The loaded column was washed with 3-5 column volumes of wash buffer (lysis buffer + 50 mM imidazole) until all non-His-tagged proteins were washed out. The target protein was then eluted with the elution buffer (lysis buffer + 200 mM imidazole). The protein elute was concentrated to around 5 ml using 15ml 10kDa filters. The concentrated protein was passed through the Econo-Pac 10DG desalting column to remove imidazole. The desalted protein was stored in the pH7 PBS buffer (50 mM KPi, 300 mM NaCl) with 10% glycerol at -80°C.

Overexpression and purification of *T.thermophilus (Tt)* ThiS thiocarboxylate (ThiS-COSH): $ThiS_{Tt}$ cloned in pTYB1 was overexpressed in *E.coli BL21(DE3)*. 1.5 L culture was grown at 37 °C in LB media until an OD_{600} of 0.6. The temperature was then reduced to 15 ^{o}C, and the culture was induced with 1 mM IPTG and continued growing at 15 ^{o}C for 12-16 hr with constant shaking (180 rpm). The culture was harvested by centrifugation and lysed by sonication on ice in 20 mM Tris, 500 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, pH 8. The lysate was passed through 0.45 μM filter and then loaded onto 20 mL chitin beads column at a rate of 0.5 mL/min and washed with 300 mL of 20 mM Tris, 500 mM NaCl, 1 mM EDTA, pH 8 at a rate of 2 mL/min. Cleavage of the ThiS protein was carried out at 4^{o}C for 48 h with 30 mL of 50 mM Na_2S to yield ThiS-COSH. The eluted ThiS-COSH was solvent-exchanged into 100 mM potassium phosphate, pH 8.0 (no reducing agent) by dialysis and stored in 50 μL 10% glycerol aliquots at -80 ^{o}C.

LRSA (90) labeling ThiS-COSH and the NH_2OH treatment: LRSA (**90**) labeling ThiS-COSH: 50 µl of ~200 µM purified ThiS-COSH was solvent-exchanged into a PBS buffer (50 mM potassium phosphate, 300 mM NaCl, 6 M urea, pH 7.0). LRSA (**90**) was added (for purified protein samples, 3 equiv, ~2 µl, of 15 mM LRSA stock in DMSO) and the sample mixtures were incubated in the dark for 15 min (or 10 min, 30 min and 1 hr for labeling time optimization) at RT. TCEP (4 µl of 250 mM TCEP stock in water) was then added, and the mixture was further incubated for 15 min in the dark at RT. The labeled samples were passed through a desalting column (Micro Bio-Spin® Columns with Bio-Gel® P-6 from Bio-rad) to reduce the free rhodamine background and loaded in a proper amount (e.g. 20 µl) for the SDS-PAGE gel analysis. The rhodamine-labeled gel band could be observed with a Typhoon FLA 9500 instrument (Ex 530 nm/Em 580 nm).

The NH_2OH treatment: The above LRSA (**90**) labeled ThiS-COSH sample was further incubated with 50 mM NH_2OH and heat at 100 °C for 20 min before loading to the SDS-PAGE gel.

***E.coli* and *Lc. lactis* growth and proteome preparation:** *E. coli BL21 (DE3)* was inoculated from a single colony into 200 LB media and incubated overnight at 37 °C with OD600 between 1-1.5. *Lc. lactis NZ3900* was inoculated from a single colony into 200 mL M17 broth supplemented with 0.5% glucose at 28 °C with agitation (120 rpm) overnight or until OD_{600} 1-1.5. The cell pellets from the two cultures were collected by centrifugation (~0.5g) and homogenized with ~0.7 ml of the lysis buffer (100 mM Tris-HCl and 300 mM NaCl, pH 8.0) in a 2 ml centrifugation tube. 1 mg chicken eggwhite lysozyme and 6.5 mg of bacterial protease inhibitor cocktail was added to it. Glass beads (~0.5 g, 0.1 mm diameter sized, BioSpec) were added (the total volume of a sample was 1.5-1.8 ml in a 2 ml centrifugation tube) to break the

cells via a Mini-Beadbeater-24 (BioSpec). Three 1-min runs were performed with 2 min cooling time on ice in between. After the cell lysis, the mixture was centrifuged at 15,000 x g for 30 min at 4 °C, and the supernatant (contained the bacterial cell proteome, around 250 µl) was collected. Then 200 µl of the supernatant was used for labeling. The proteome in the cell extract was precipitated out in methanol/chloroform and redissolved in the PBS buffer for labeling.

LRSA (90) labeling E. Coli proteome and the NH_2OH treatment: 200 µl E. coli proteome (from 200 ml cell culture) was solvent-exchanged into a PBS buffer (50 mM potassium phosphate, 300 mM NaCl, 6 M urea, pH 7.0). LRSA (**90**) (10 µl of 15 mM stock in DMSO) was added, and the sample mixtures were incubated in the dark for 10min, 20 min, 30 min or 1hr at RT. TCEP (6 µl of 250 mM stock in water) was then added, and the mixture was further incubated for 15 min in the dark at RT. The labeled samples were passed through a desalting column. 50 mM NH_2OH (**99**) was then added to the mixture and heated at 100 °C for 15 min, 30 min, and 45 min. The samples were loaded in a proper amount (e.g. 20 µl) for the SDS-PAGE gel analysis. The LRSA-labeled gel band could be observed with a Typhoon FLA 9500 instrument (Ex 530 nm/Em 580 nm).

LRSA (90) labeling *Lc. lactis* proteome under optimized conditions: 200 µl *Lc. lactis* proteome (from 200 ml cell culture) was solvent-exchanged into a PBS buffer (50 mM potassium phosphate, 300 mM NaCl, 6 M urea, pH 7.0) and incubated for 15 min at RT. LRSA (**90**) (10 µl of 15 mM stock in DMSO) was added, and the sample mixtures were incubated in the dark for 15 min at RT. TCEP (6 µl of 250 mM stock in water) was then added, and the mixture was further incubated for 15 min in the dark at RT. The labeled samples were passed through a desalting column and loaded in a proper amount (e.g. 20 µl) for the SDS-PAGE gel analysis. The

LRSA-labeled gel band could be observed with a Typhoon FLA 9500 instrument (Ex 530 nm/Em 580 nm).

HPLC parameters: The HPLC model was an Agilent 1200 or 1260 with a quaternary pump. The system included a diode array UV-Vis detector, and eluted compounds were detected using absorbance at 254, 280, 260, 567, 220 and 340 nm. The analysis was performed on a ZORBAX Eclipse XDB-C18 column (4.6 x 150 mm, 5 μm particles, Agilent Technologies). Data were processed using ChemStation version B.04.01 SP1 (Agilent Technologies).

HPLC conditions:

A- Water

B- 10mM Ammonium acetate, pH 6.6

C- Methanol

HPLC method: (Flow rate = 1 ml/min)

0 min: 100% A; 7 min: 90% A, 10% B; 12 min: 25% A, 60% B, 15% C; 16 min: 25% A, 10% B, 65% C; 18 min: 100% B; 30 min: 100% B.

LC-MS parameters:

LC-ESI-TOF-MS was performed using an Agilent 1260 HPLC system equipped with a binary pump and a 1200 series diode array detector followed by a MicroTOF-Q II mass spectrometer (Bruker Daltonics) using an ESI source either in positive mode or negative mode. The analysis was performed on a Poroshell 120 EC-C18 column (3.0 x 100 mm, 2.7 μm particles, Agilent Technologies).

LC conditions:

A: 5 mM ammonium acetate buffer, pH 6.6

B: 75% Methanol and 25% Water

LC method: (Flow rate = 0.4 ml/min, for both the positive and negative mode on MS)

0 min: 100% A, 0% B; 2 min: 100% A, 0% B; 12 min: 25% A, 75% B; 17 min: 25% A, 75% B;

18.5 min: 100% A, 0% B; 30 min: 100% A, 0% B.

5. LABELING PROTEIN THIOCARBOXYLATES INVOLVED IN THE BIOSYNTHESIS OF THE NAD DERIVED PINCER COFACTOR

5.1 Protein thiocarboxylates involved in the LarA biosynthesis pathway

Many enzymes need cofactors to reconstitute their activity. Some cofactors are covalently bound to their partner enzymes in a post-translational modification fasion. In 2015, a novel NAD derived Ni-pincer cofactor was found covalently bound to the lactate racemase, LarA, in *Lactobacillus plantarum* (Figure 5.1).[27] Interestingly, this cofactor has a thiocarboxylate moiety. Thus, LarA can be regarded as a protein with thiocarboxylation PTM, i.e. a protein thiocarboxylate. Protein thiocarboxylation used to be considered only taking place at the C-terminal glycine carboxylate of the ThiS-like sulfur carrier protein and play a role in sulfur transfer and activation. The cofactor-attached LarA has then become the second example of protein thiocarboxylate with no structural or functional similarity to ThiS. The molecular weight of LarA is around 48 kDa, which is much higher than ThiS (< 10 kDa). Its dithiocarboxylate moiety is buried inside the binding pocket and functions as the chelating ligand for a nickel ion.

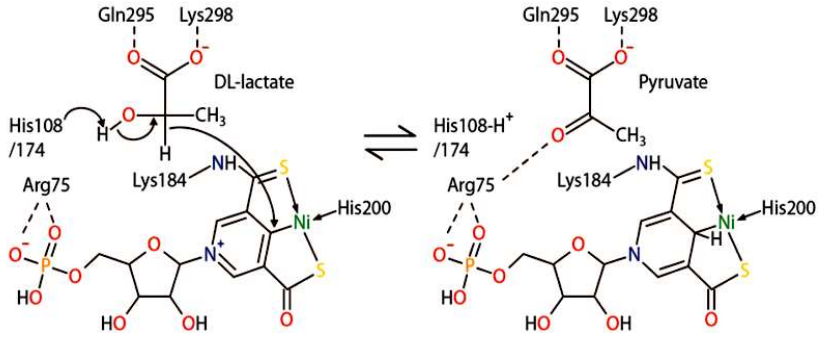


Figure 5.1: LarA catalyzed reaction. The NAD derived thiocarboxylate cofactor is covalently bound to the Lys184 of LarA.

Lactobacillus plantarum (*L. plantarum*, or *Lp*) needs both L-lactate and D-lactate for growth. LarA is the lactate racemase responsible for the L/D-lactate conversion.[87] Crystal structure and mass spectrometry data have shown that the active form of LarA has a special NAD derived thiocarboxylate pincer cofactor, nickel pyridinium-3,5-bisthiocarboxylic acid mononucleotide (**16**), covalently attached (Figure 5.2).[27] In the same larABCDE gene operon, LarB, LarE, and LarC are sufficient for the biosynthesis of **16** from nicotinic acid adenine dinucleotide (NaAD, **8**), while LarD is not required.[88] LarB first catalyzes the carboxylation of **8** and forms pyridinium-3,5-biscarboxylic acid mononucleotide (P2CMN, **9**). Two equivalents of ATP dependent LarE then convert **9** sequentially into two thiocarboxylate small metabolites, pyridinium-3-carboxy-5-thiocarboxylic acid mononucleotide (PCTMN, **12**) and pyridinium-3,5-bisthiocarboxylic acid mononucleotide (P2TMN, **15**). In between, the active cysteine176 of the second equivalent of LarE can interact with the activated species, **13**, covalently and form the labile protein thiocarboxylate, LarE-COSH (**14**). **15** is then released from **14** through a general base-catalyzed cysteine desulfuration.[89] LarC is a CTP-dependent nickelochelatase which incorporates a nickel (II) ion into **15** to make **16**. In the active LarA (**17**), **16** will form a thioamide bond with the Lys184 of LarA. **17** can readily lose its nickel ion non-enzymatically and become the much more stable protein thiocarboxylate covalent adduct, LarA-COSH (**18**). Apart from the core enzymes LarA, LarB, LarE, and LarC, there are some other protein members coded in the lar gene cluster (Figure 5.3), such as a nickel transporter (Lar(MN)QO), a lactic acid channel (LarD) and a transcriptional regulator (LarR).[87]

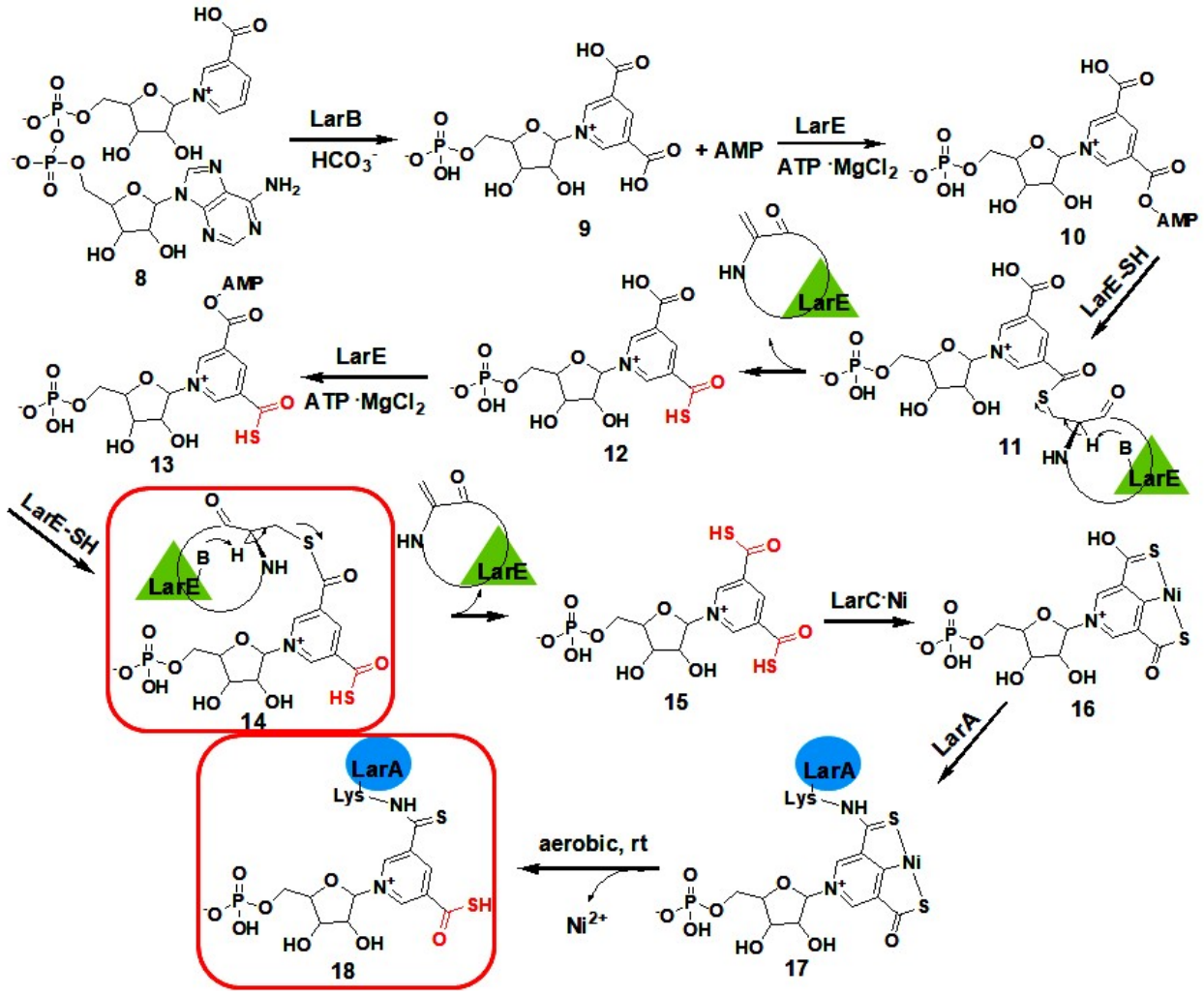


Figure 5.2: The two protein thiocarboxylates in the LarA cofactor biosynthesis pathway, LarA-COSH (**18**) and LarE-COSH (**14**) (red boxed), were selectively labeled with the sulfonyl azids "click reagents".

Figure 5.3: The gene cluster of the LarA activation pathway. LarA, LarB, LarE, and LarC are required for LarA cofactor biosynthesis. Lar(MN)QO complex is a nickel transporter, LarD is a lactic acid channel, and LarR is a transcriptional regulator. LarD is not required for LarA cofactor biosynthesis.

5.2 Labeling the two protein thiocarboxylates in the LarA cofactor biosynthesis pathway using sulfonyl azide reagents (LRSA, 90 and BiotinSA, 91, see Figure 5.4)

Bioinformatics study has shown that the LarA cofactor (**16**, Figure 5.2) biosynthesis genes (*larA, larB, larC and larE*) are widely distributed in bacterial species.[87] In many strains, *larA* or other *lar* genes are missing from the LarA cofactor biosynthesis gene clusters, which indicates the possibility of **16** to serve other enzymes to achieve new functions rather than lactate racemization. The biosynthesis of cofactor **16** requires the coexpression of *larABCE,* and it is time-consuming to clone these genes together and separately (as controls) for studying the function of the **16**-involved non-LarA systems. It may take more time to optimize the overexpression and purification conditions. It is desirable to develop a highly selective and sensitive method to detect the low concentration of **16**-attached proteins directly from the bacterial proteomes before further biochemistry characterisation. The thiocarboxylate group of **16** becomes a potential target for sulfonyl azide reagents labeling. In Chapter 4, I optimized the labeling condition by minimizing the non-specific lysine-sulfonyl azide interaction. In this chapter, I applied this strategy to label the two protein thiocarboxylates, LarE-COSH (**14**) and LarA-COSH (**18**) (Figure 5.2, in red boxes), involved in the LarA activation/degradation in the bacterial proteomes. With the structural and functional distinct properties, we investigated if these sulfonyl azide reagents can label non-ThiS-like protein thiocarboxylates selectively and sensitively. Both Lissamine Rhodamine B Sulfonyl Azide (LRSA, **90**) and biotin-4-carboxybenzene sulfonyl azide (BiotinSA, **91**) were used in this labeling study (Figure 5.4).

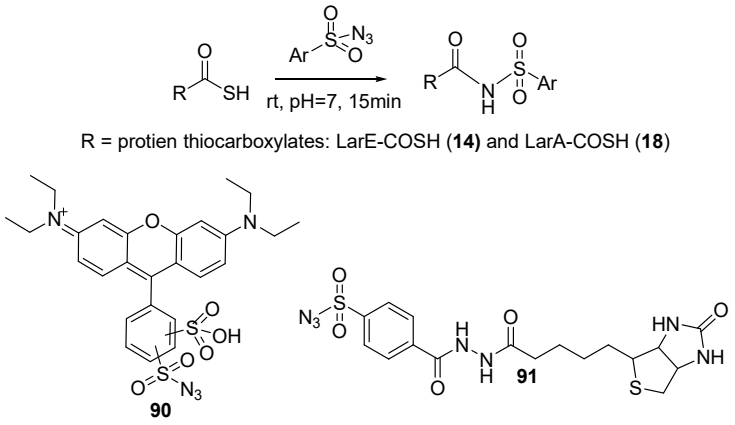


Figure 5.4: The critical trapping "click reaction" between aryl sulfonyl azides (**90 and 91**) and the protein thiocarboxylates: LarE-COSH (**14**) and LarA-COSH (**18**). **90**: Lissamine Rhodamine B Sulfonyl Azide (LRSA); **91**: biotin-4-carboxybenzene sulfonyl azide (BiotinSA).

5.3 Results and discussions

5.3.1 Labeling LarA thiocarboxylate (LarA-COSH, 18) using Lissamine Rhodamine B Sulfonyl Azide (LRSA, 90)

The active LarA from *L. plantarum* (**17**) was purified through the overexpression of the complete larABCDE operon (plasmid pGIR100) in *Lactococcus lactis* (*Lc. lactis*) *NZ3900* strain with a StrepII-tag fused at its C-terminal[27]. Under ambient conditions without reducing reagent, most **17** rapidly lost the nickel ion and converted to **18**, a better protein thiocarboxylate nucleophile than **17** to interact with LRSA (**90**). **18** was characterized by the whole protein mass (Figure 5.5). Lissamine Rhodamine B Sulfonyl Azide (LRSA, **90**) was synthesized following the reported protocol[21]. In Figures 5.6 and 5.7, SDS-PAGE protein gel showed that **18** could be labeled by **90** with approximately 20 pmol protein sensitivity, which is comparable to the reported ~10 pmol sensitivity of **90** labeling ThiS-like protein thiocarboxylates[21]. The sensitivity

was measured by adding varying concentrations of purified **18** into the set amount of *Lc. Lactis* NZ3900 proteome (no *lar* operon was found in *Lc. Lactis NZ3900* genome).

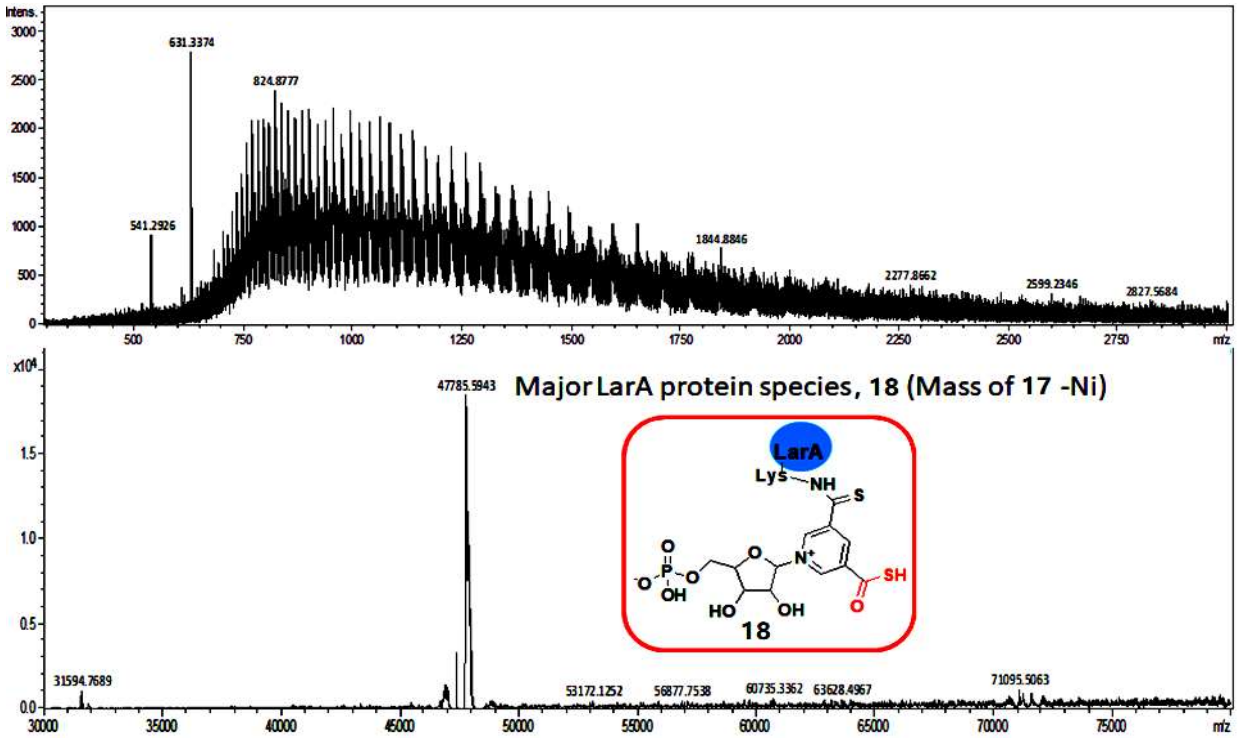


Figure 5.5: Whole protein mass of the stable LarA-COSH (**18**) in aerobic condition.

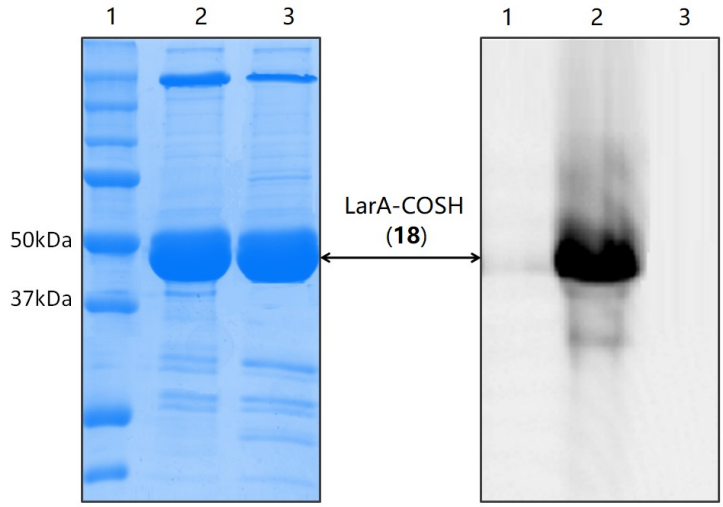


Figure 5.6: The SDS-PAGE analysis of the LRSA (**90**) labeled the purified LarA-COSH (**18**) at 48 kDa. Left: the Coomassie blue stain; Right: the fluorescence image at Ex/Em 530/580 nm (rhodamine fluorescence range). Lane 1: the protein ladder; lane2: the purified LarA-COSH (**18**) at 48 kDa; lane 3: the purified LarA-COSH (**18**) at 48 kDa labeled by LRSA (**90**) for 15 min in the dark at RT and quenched by TCEP.

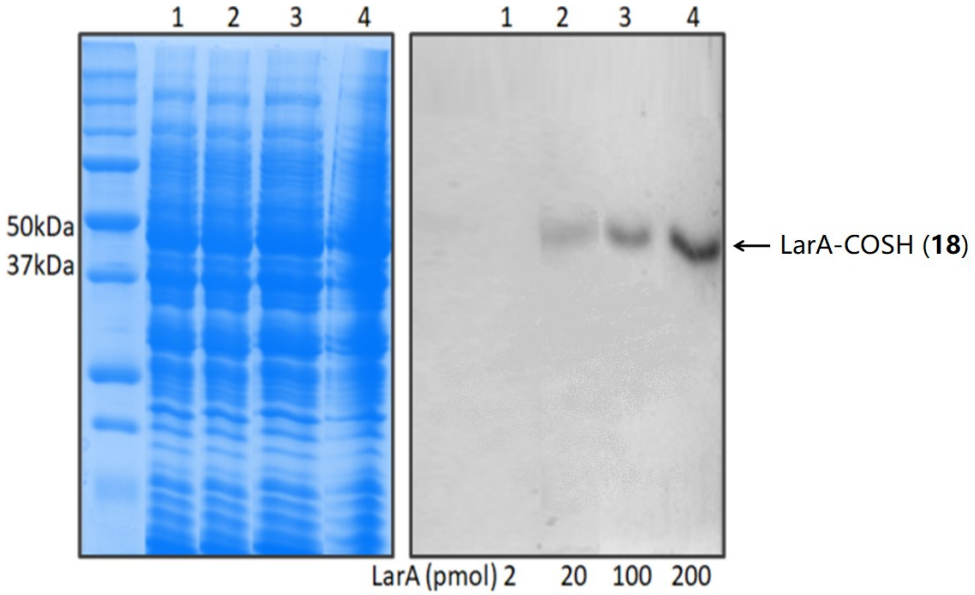


Figure 5.7: The SDS-PAGE analysis of LarA-COSH (**18**) labeling limit with LRSA (**90**). Lane 1-4 were all loaded with 15µl *Lc. lactis NZ3900* proteome, and sequentially added 2pmol, 20pmol, 100pmol, and 200pmol purified **18** to lane 1 to 4. All samples were labeled by the same excess amount of **90** (2 mM) before gel loading. Left: Coomassie blue staining; right: the fluorescence image at Ex/Em 530/580 nm.

5.3.2 Labeling LarE protein thiocarboxylate (LarE-COSH, 14) using Lissamine Rhodamine B Sulfonyl Azide (LRSA, 90)

To obtain the labile LarE-COSH (**14**), LarB and LarE were overexpressed and purified separately, and then they were performed with a coupled enzymatic reaction using **8** as the initial substrate (Figure 5.2). The freshly made **14** could be labeled explicitly by **90** while LarB did not get labeled due to the lack of thiocarboxylation modification (Figure 5.8). **14** was reported to undergo general base catalysis of cofactor-attached cysteine desulfuration and fully release the small thiocarboxylate, **15**, within 2 hours[89]. The thiocarboxylate-releasing process was further supported by placing fresh **14** samples on the bench for various time intervals and before **90** labeling (Figure 5.9). LarE could not be labeled anymore once it lost its thiocarboxylate group. Because of the instability of **14**, the labeling sensitivity could not be accurately measured.

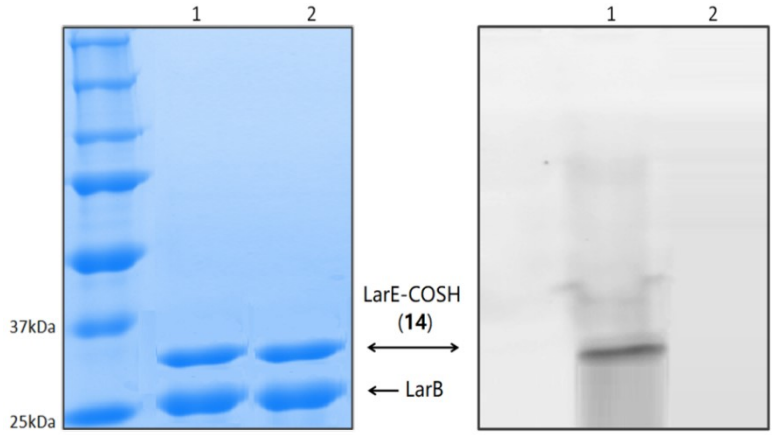
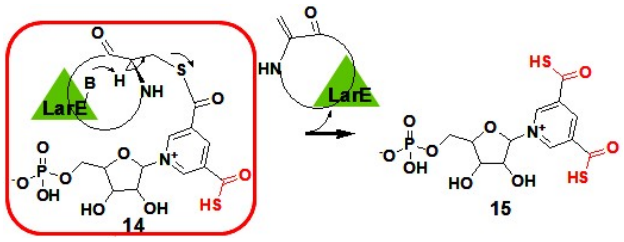


Figure 5.8: The SDS-PAGE analysis of LarE-COSH (**14**, 32 kDa) generated from purified LarB/LarE coupled reaction with **8** was labeled by LRSA (**90**). LarB (27 kDa) was not labeled. Lane 1: fresh-made LarE-COSH (**14**) labeled with LRSA (**90**); lane 2: fresh-made LarE-COSH (**14**) with no LRSA (**90**). Left: Coomassie blue staining; right: the fluorescence image at Ex/Em 530/580 nm.

A)



B)

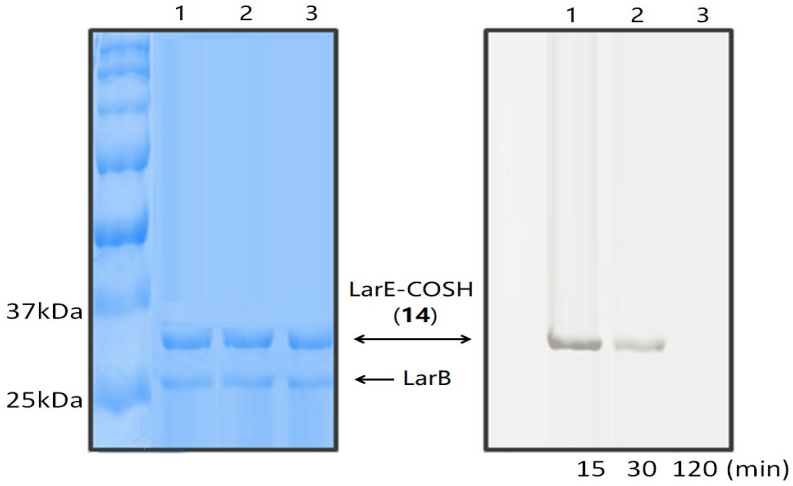


Figure 5.9: A) The cartoon of LarE-COSH (**14**) releasing the small thiocarboxylate cofactor, **15**. B) The SDS-PAGE analysis of LarE-COSH (**14**) fully released the small thiocarboxylate cofactor, **15**, within 2 h. Fresh LarE-COSH (**14**) was seated on the bench at RT for 15min (lane 1), 30min (lane 2) and 2h (lane 3) and then labeled with LRSA (**90**). Left: Coomassie blue staining; right: the fluorescence image at Ex/Em 530/580 nm.

5.3.3 Trapping $LarA_{Lp}$-COSH (18) and LarE-COSH (14) in various bacterial proteomes using LRSA (90) and biotin-4-carboxybenzene sulfonyl azide (BiotinSA, 91)

LRSA (**90**) was further tested to trap LarE-COSH (**14**) and $LarA_{Lp}$-COSH (**18**) selectively in $LarA_{Lp}$BCDE overexpressed *Lc. lactis* proteome (plasmid pGIR100). Only two bands corresponding to $LarA_{Lp}$-COSH (**18**) and LarE-COSH (**14**) were found at

ex530nm/em580nm (the typical lissamine rhodamine B fluorescent range) with molecular weights around 48kDa and 32kDa (Figure 5.10, lane 1). $LarA_{Tt}$ from *Thermoanaerobacterium thermosaccharolyticum* (Tt) is an ortholog which shares 53% sequence identity with $LarA_{Lp}$. The crystal structure of $LarA_{Tt}$ showed that it did not form the covalent bond with the thiocarboxylate pincer cofactor **16**[27]. Thus $LarA_{Tt}$ should not be labeled by **90**. As a control, we labeled $LarA_{Tt}$BCDE overexpressed *Lc. lactis* proteome (plasmid pGIR182) with **90** and found only LarE-COSH (**14**) but no $LarA_{Tt}$ band shown on the gel (Figure 5.10, lane 2).

The *$larA_{Lp}$BCDE* operon was coexpressed in *Lc. lactis* (pGIR100) to generate the active form of $LarA_{Lp}$, **17** (Figure 5.2 and 5.3). Single deletion of *larA*, *larB*, *larC* or *larE* gene may interrupt the formation of LarA-COSH (**18**) or LarE-COSH (**14**) and cause the failure of LarA/LarE labeling in the proteomes. Single deletion of LarD resulted in labeling of both LarA and LarE since LarD is a lactic acid channel and is not required for LarA activation (Figure 5.11, lane 1). The LarA, LarC2 (the essential C-terminal of LarC) or LarC deleted strains could only produce LarE-COSH (**14**) but not LarA-COSH (**18**) (Figure 5.11, lane 2-4). Therefore, only the **14** fluorescent band was observed. Deletion of *larB* or *larE* from the *larABCDE* caused no production of **18** or **14**. Thus neither LarA nor LarE could be labeled by **90** (Figure 5.11, lane 5 and 6). This set of controls further supported that **90** could selectively label LarA-COSH (**18**) and LarE-COSH (**14**) through a covalently bound thiocarboxylate cofactor moiety.

Furthermore, **90** could also successfully label LarA-COSH (**18**) in the wild-type *L. plantarum* proteome. LarE was not detected by **90** labeling, probably due to the instability of LarE-COSH (**14**), as well as the different expression and thiocarboxylation levels between LarA and LarE in the wild-type proteome (Figure 5.12).

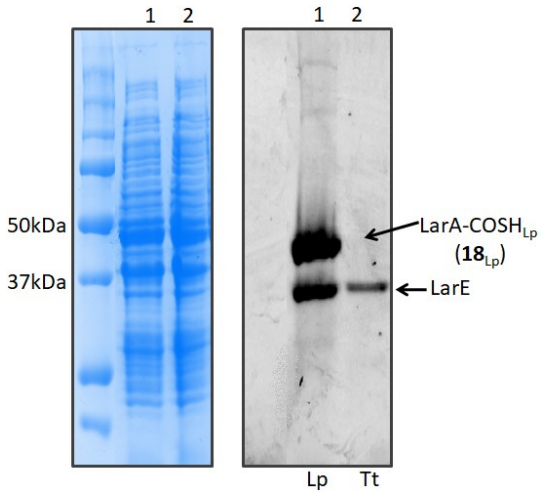


Figure 5.10: The SDS-PAGE analysis of LRSA (**90**) labeling both LarA$_{Lp}$-COSH (**18**$_{Lp}$) and LarE-COSH (**14**) in LarA$_{Lp}$BCDE overexpressed *Lc. lactis* proteome (lane 1). In LarA$_{Tt}$BCDE overexpressed *Lc. lactis* proteome, only LarE-COSH (**14**) was labeled by **90** (lane 2). LarA$_{Tt}$ was not labeled since its thiocarboxylate cofactor did not covalently bind to the protein. Lp: *Lactobacillus plantarum.* Tt: *Thermoanaerobacterium thermosaccharolyticum.* Left: Coomassie blue staining; right: the fluorescence image at Ex/Em 530/580 nm.

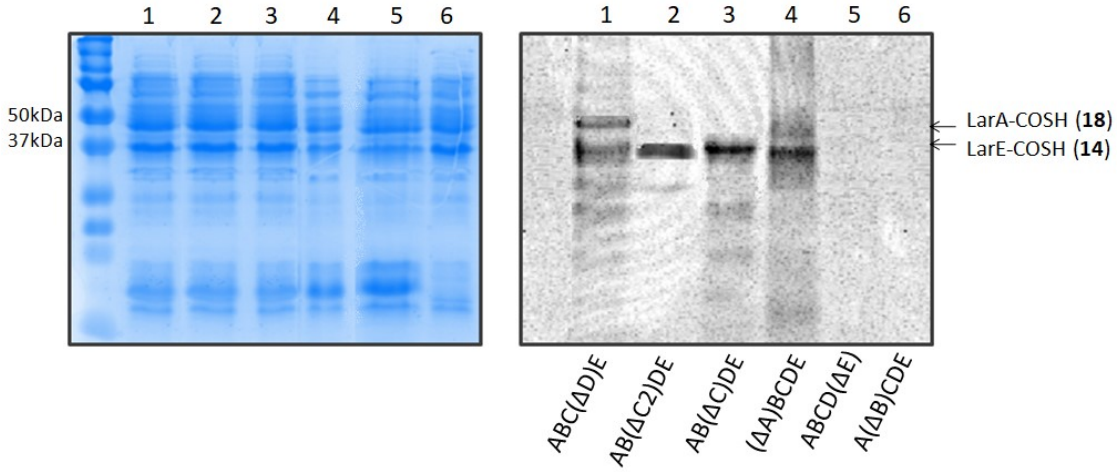


Figure 5.11: The SDS-PAGE analysis of LRSA (**90**) labeling LarA-COSH (**18**) and/or LarE-COSH (**14**) from *Lc. lactis* proteomes with lar operon variants (single operon deletion, Δ) overexpressed. Lane 1, LarABC(ΔD)E (produced **14** and **18**); lane 2, LarAB(ΔC2)DE, LarC2 is the essential C-terminal of LarC (produced only **14**); lane 3, LarAB(ΔC)DE (produced only **14**); lane 4, Lar(ΔA)BCDE (produced only **14**); lane 5, LarABCD(ΔE) (produced neither **14** nor **18**); lane 6, LarA(ΔB)CDE (produced neither **14** nor **18**). Left: Coomassie blue staining; right: the fluorescence image at Ex/Em 530/580 nm.

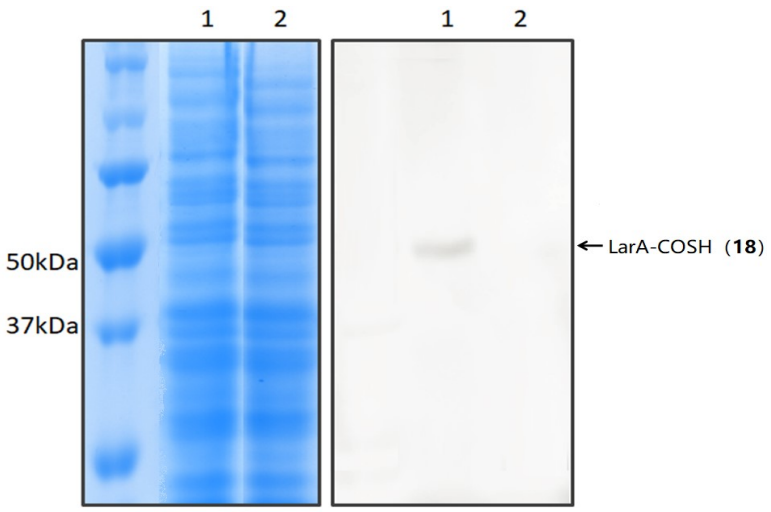


Figure 5.12: The SDS-PAGE analysis of LRSA (**90**) labeling LarA-COSH (**18**) in wild-type *L. plantarum* (Lp_{WT}) proteome induced with L-lactate sodium salt. Lane 1: the Lp_{WT} proteome labeled with LRSA (**90**); lane 2: the Lp_{WT} proteome**.** Left: Coomassie blue staining; right: the fluorescence image at Ex/Em 530/580 nm.

The $LarA_{Lp}BCDE$ (plasmid pGIR100) overexpressed *Lc. lactis* proteome was also labeled with biotin-4-carboxybenzene sulfonyl azide (BiotinSA, **91**) for the protein thiocarboxylates enrichment and LC-MS/MS peptide characterization. Before loading onto the streptavidin agarose resin for biotinylated LarA-COSH (**18**) enrichment, a small portion of the labeled proteome was directly analyzed using the streptavidin-R-phycoerythrin (PE) staining western blot technique. Only one significant fluorescent band (Ex 530nm/Em 580nm) could be observed at ~50 kDa (Figure 5.13, lane 2). After the streptavidin resin enrichment, the biotinylated proteins were eluted and analyzed by SDS-PAGE gel (Coomassie blue staining), western blot (streptavidin-R-phycoerythrin (PE) staining), and trypsin digested peptide LC-MS/MS techniques. The protein band at ~50 kDa was significantly enriched (Figure 5.13, lane 1 & 3). The major eluted protein was identified as $LarA_{Lp}$ with 91% peptide coverage by the

peptide LC-MS/MS characterization (Figure 5.14), though the biotinylated cofactor attached peptide was not detected. LarE-COSH was not labeled due to its instability (these methods required a longer time than direct fluorescence labeling assays with **90**).

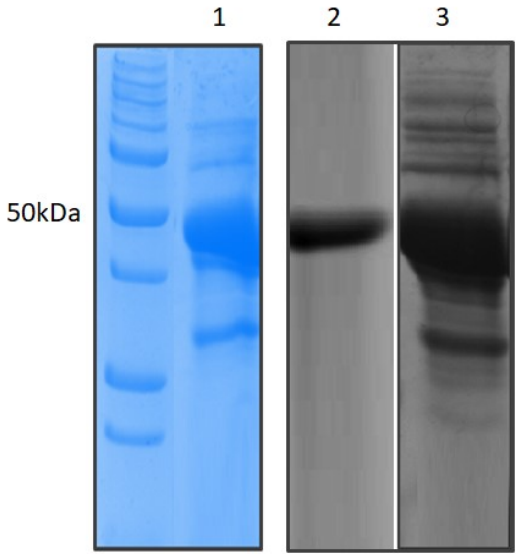


Figure 5.13: BiotinSA (**91**) labeled LarA-COSH (**18**) in $LarA_{Lp}BCDE$ overexpressed *Lc. lactis* proteome. Lane 1 and 3: $LarA_{Lp}BCDE$ overexpressed *Lc. lactis* proteome + **91**, passed through the streptavidin agarose resin, washed, eluted (in 10 mM EDTA, 95% formamide, 5 mM free biotin at 70 °C for 5 min), separated in an SDS-PAGE gel and visualized by both Coomassie blue staining (lane 1) and a western blot assay (streptavidin-R-phycoerythrin (PE) staining, lane 3); lane 2: $LarA_{Lp}BCDE$ overexpressed *Lc. lactis* proteome + **91**, directly separated by an SDS-PAGE gel and visualized by a western blot assay (streptavidin-PE staining) without the streptavidin agarose resin enrichment. Lane 1: Coomassie blue staining; lane 2 & 3: the fluorescence image at Ex 530nm/Em 580nm (the streptavidin-PE staining fluorescence range).

Protein View: gi|887492742

Chain A, A Tethered Niacin-derived Pincer Complex With A Nickel-carbon Bond In Lactate Racemase

Database:	NCBInr
Score:	227
Expect:	1.3e-15
Nominal mass (M_r):	47534
Calculated pI:	5.60
Taxonomy:	Lactobacillus plantarum WCFS1

Enzyme:	Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Deamidated (NQ)
Mass values searched:	249
Mass values matched:	119

Protein sequence coverage: 91%

Matched peptides shown in ***bold red***.

```
  1 SVAIDLPYDK RTITAQIDDE NYAGKLVSQA ATYHNKLSEQ ETVEKSLDNP
 51 IGSDKLEELA RGKHNIVIIS SDHTRPVPSH IITPILLRRL RSVAPDARIR
101 ILVATGFHRP STHEELVNKY GEDIVNNEEI VMHVSTDDSS MVKIGQLPSG
151 GDCIINKVAA EADLLISEGF IESHFFAGFS GGRKSVLPGI ASYKTIMANH
201 SGEFINSPKA RTGNLMHNSI HKDMVYAART AKLAFIINVV LDEDKKIIGS
251 FAGDMEAAHK VGCDFVKELS SVPAIDCDIA ISTNGGYPLD QNIYQAVKGM
301 TAAEATNKEG GTIIMVAGAR DGHGGEGFYH NLADVDDPKE FLDQAINTPR
351 LKTIPDQWTA QIFARILVHH HVIFVSDLVD PDLITNMHME LAKTLDEAME
401 KAYAREGQAA KVTVIPDGLG VIVKASWSHP QFEK
```

Figure 5.14: LC-MS/MS peptide coverage (in red) of LarA enriched from $LarA_{Lp}BCDE$ overexpressed *Lc. lactis* proteome (the same sample as loaded in figure 73, lane 1). $LarA_{Lp}BCDE$ overexpressed *Lc. lactis* proteome was labeled with BiotinSA (**91**), desalted, passed through the streptavidin agarose resin, washed and eluted the enriched biotinylated LarA-COSH (**18**). The eluted protein was trypsin digested and analyzed by LC-MS/MS.

5.4 Conclusion

The two protein thiocarboxylates, LarA-COSH (**14**) and LarE-COSH (**18**) were selectively labeled by LRSA (**90**) in both the purified form and the bacterial proteome mixtures (both the overexpression *Lc. lactis* strain and the wild-type *L. plantarum* strain). LarA-COSH (**14**) had the labeling limit of 20 pmol for the SDS-PAGE gel fluorescence imaging. LarE-COSH (**18**) was not stable and lost its thiocarboxylate cofactor within 2 hr. BiotinSA (**91**) labeled LarA-COSH (**18**)

in the overexpression strain's proteome was observed by western blot (streptavidin-R-phycoerythrin (PE)) and was enriched for LC-MS/MS peptide identification with high peptide coverage.

5.5 Materials and methods

Materials: All chemicals were purchased from Sigma-Aldrich unless otherwise mentioned. LB-Lennox broth was from EMD Millipore. Kanamycin and IPTG were obtained from Lab Scientific Inc. Chloramphenicol was from Fisher Scientific. QIAprep Spin Miniprep Kit was obtained from QIAGEN. Gravity flow Strep-Tactin® Superflow® high capacity columns were purchased from IBA GmbH. D-chloroform, D_6-DMSO, D_4-methanol were purchased from Cambridge Isotope Laboratories. 2.5L baffled ultra yield flasks for overexpressing proteins were purchased from Thomson Instrument Company. The streptavidin resin was from GenScript (Piscataway, NJ). The streptavidin-R-phycoerythrin (PE) was from QIAGEN (Valencia, CA). The polyvinylidene difluoride (PVDF) membrane was from Promega (Madison, WI). The fluorescence images were scanned with a Typhoon FLA 9500 instrument from GE Healthcare Biosciences (Piscataway, NJ). The OD_{600} and other UV-vis absorptions were measured using a Varian Cary 300 Bio UV-Visible Spectrophotometer (Palo Alto, CA). The plasmids and strains used in this project were kindly provided by Dr. Robert P. Hausinger's lab (Table 5.1).

Table 5.1: Plasmids and strains used in this study (provided from Dr. Hausinger's lab)

Plasmids	Strains	Characteristics	Functions
pGIR100	*Lactococcus lactis* (*Lc. lactis*) *NZ3900*	Cm^r; pNZ8048 with a 4.92-kb insert containing the translational fusion between P_{nisA} and *$larA_{Lp}$-larE*	$LarA_{Lp}$-LarE overexpression for proteome labeling
pGIR112	*Lc. lactis NZ3900*	Cm^r; pGIR100 with the Strep-tag fused at the C-terminus of $LarA_{Lp}$.	$LarA_{Lp}$ purification
pGIR200	*Lc. lactis NZ3900*	Cm^r; pGIR100 with an in-frame deletion of *larA*	Lar(ΔA)BCDE overexpression (proteome labeling control)
pGIR300	*Lc. lactis NZ3900*	Cm^r; pGIR100 with an in-frame deletion of *larB*	LarA(ΔB) CDE overexpression (proteome labeling control)
pGIR400	*Lc. lactis NZ3900*	Cm^r; pGIR100 with an in-frame deletion of *larC2*	LarABC1(ΔC2)DE overexpression (proteome labeling control)
pGIR500	*Lc. lactis NZ3900*	Cm^r; pGIR100 with an in-frame deletion of *larC*	LarAB(ΔC)DE overexpression (proteome labeling control)
pGIR600	*Lc. Lactis NZ3900*	Cm^r; pGIR100 with an in-frame deletion of *larD*	LarABC(ΔD)E overexpression (proteome labeling control)
pGIR700	*Lc. lactis NZ3900*	Cm^r; pGIR100 with an in-frame deletion of *larE*	LarABCD(ΔE) overexpression (proteome labeling control)

pGIR182	*Lc. lactis NZ3900*	Cm^r; pGIR112 with *larA* of *L. plantarum* ($larA_{Lp}$) replaced by *larA* of *T. thermosaccharolyticum* ($larA_{Tt}$)	$LarA_{Tt}BCDE$ overexpression (proteome labeling control)
pGIR026	*E. coli DH10B*	Ap^r Gm^r; pBADHisA with a 0.77-kb insert containing the translational fusion between P*nisA* and *larB*-strep-tag fusion (at the 3′-end of the *larB* ORF)	LarB purification
pGIR076	*E. coli* arctic express (DE3)	Ap^r Gm^r; pBADHisA with a 0.89-kb insert containing the translational fusion between P_{nisA} and larE-strep-tag fusion	LarE purification
pGIR051	*Lc. lactis NZ3900*	Cm^r; pNZ8048 with a 1.30-kb insert after P_{nisA} containing DNA encoding the StrepII-tag translationally fused to *larC* with a 1 bp insertion between *larC1* and *larC2*	LarC purification
	Lactobacillus plantarum (*L. plantarum*, or *Lp*) NCIMB8826	Wild-type	Native strain with larA-larE in the genome, for proteome labeling

Table 5.1: continued

Growth conditions: *L. plantarum* was grown in De Man–Rogosa–Sharpe (MRS) broth at 28 °C without shaking. For induction of Lar activity in *L. plantarum*, L-lactate sodium salt was added at a concentration of 200mM during the mid-log phase (OD_{600} = 0.6-0.7), and the cells were collected 4 h later.

Lc. lactis was grown in M17 broth supplemented with 0.5% glucose at 28 °C at 120 rpm. Chloramphenicol (10 mg/ml) and $NiCl_2$ (1mM) were added to the media. For the induction of genes under control of the nisA promoter, nisin A was added during the early exponential phase (OD_{600} = 0.3–0.4) at a concentration of 1 $\mu g \cdot L^{-1}$, and the cells were collected after 4 h.

E. coli was grown with agitation (200 rpm) at 37 °C in LB containing ampicillin (100 $mg \cdot L^{-1}$) and gentamicin (25 $mg.L^{-1}$). For the pBAD expression system, when the OD600 = 0.5-0.8, the cells were induced with 0.2% L-arabinose. After induction, *E. coli DH10B* cells continually grew overnight at 25 °C before harvesting, while *E. coli* arctic express cells were incubated for 24-26 h at 13 °C.

The cells were harvested by centrifugation at 9000 rpm for 10 min at 4 °C, and the pellet was stored at -80 °C.

Bacterial cell extract and proteome preparation: The cell pellet (~0.5 g from 200 ml culture, stored at -80 °C) was homogenized with ~0.7 ml of the lysis buffer (100 mM Tris-HCl and 300 mM NaCl, pH 8.0) in a 2 ml centrifugation tube. 1 mg chicken eggwhite lysozyme and 6.5 mg of bacterial protease inhibitor cocktail was added to it. Glass beads (~0.5 g, 0.1 mm diameter sized, BioSpec) were added (the total volume of a sample was 1.5-1.8 ml in a 2 ml centrifugation tube) to break the cells via a Mini-Beadbeater-24 (BioSpec). Three 1-min runs were performed with 2 min cooling time on ice in between. After the cell lysis, the mixture was centrifuged at 15,000 x g for 30 min at 4 °C, and the supernatant (contained the bacterial cell

proteome, around 250 µl) was collected. Then 200 µl of the supernatant was used for labeling. The proteome in the cell extract was precipitated with methanol/chloroform and redissolved in the PBS buffer for labeling. For the purification of *lar* proteins, cells from 2 L bacterial cultures were collected and resuspended in 7-10 ml lysis buffer. Such cell resuspension was put into 2ml centrifugation tubes in 0.7-1 ml aliquots (~10 tubes in total) and the cells were lysed following the above protocol with a Mini-Beadbeater-24 (BioSpec). The supernatants (from the 10 tubes) were combined for the following protein purification.

StrepII-tagged proteins (LarA, LarB and LarE) purification: 30 ml of supernatant was added to a Gravity flow Strep-Tactin® Superflow® high capacity column (1 ml). The column was washed using 5 ml of washing buffer (100 mM Tris-HCl, 300 mM NaCl, 1 mM DTT, pH 7.5-8, for LarA purification, 50 µM Na_2SO_3 was added). The cell lysate was passed through the 0.45 µm filter before being loaded onto the column, followed by 10 ml washing buffer. Then the target protein was eluted with 4x1 ml of elution buffer (washing buffer + 2.5 mM desthiobiotin, pH 8). The column was regenerated according to the manufacturer's protocol and could be reused for 4-5 times. The protein yield ranged from 5-15 mg per liter of culture. The proteins were concentrated using Amicon® Ultra 10 kD centrifugal filters (Millipore, Billerica, Massachusetts) to the appropriate concentrations. The desthiobiotin was removed using the Econo-Pac 10DG desalting column. The proteins were used immediately or stored at 4 °C for no longer than one day. The protein was solvent-exchanged into 50 mM ammonium acetate pH 6.6 buffer for the whole protein LC-MS analysis (final injection: 50 µl of 200 µM protein samples).

Protein activity assays: LarA: The LarA activity was measured by the D- to L-lactate or L- to D-lactate conversion. The cell extract or purified active form of LarA protein was incubated with 0.1-5 mM D- or L-lactate in 60 mM Tris-maleate buffer at pH 6.0 (TM buffer) at 35 °C

($LarA_{Lp}$) or 50 °C ($LarA_{Tt}$) for 10 min. The reaction was quenched by heat-denaturing the protein at 90 °C for 10 min. The lactate conversion was measured by a D-lactic acid/L-lactic acid commercial kit (R-Biopharm, Germany). The L-lactate dehydrogenase and D-lactate dehydrogenase provided in the kit could specifically oxidize L-lactate or D-lactate to pyruvate using NAD. NADH was formed, and its characteristic UV absorbance at 340 nm could be detected using a Varian Cary 300 Bio UV-Visible Spectrophotometer (Palo Alto, CA).

LarB: The reaction mixture contained LarB (5 μM), NaAD (**8**) (0.5 mM), $NaHCO_3$ (10 mM), and $MgCl_2$ (4 mM) in Tris·HCl buffer (100 mM, pH 8). It was incubated for overnight at RT, and the reaction was quenched by heat treatment at 90 °C for10 min. The protein was filtered out through a 10kDa 1.5ml centrifugal filter, and the filtrate was analyzed by LC-MS to check the formation of the carboxylated product **9** ([M+1] = 380.0838).

LarE: The product **9** from the previous LarB carboxylation was reacted with LarE (100 μM), ATP (5 mM), and $MgCl_2$ (20 mM) in Tris·HCl buffer (100 mM, pH 7) with a final volume of 250 μl. The reaction was quenched by heat denaturing (60 °C, 5 min) after a 2 hr incubation at RT, and the protein was removed from a 10kDa 1.5ml centrifugal filter. The filtrate was concentrated to 50 μl and analyzed by LC-MS to determine the formation of the thiocarboxylated product **12** or **15** ([M+1] = 396.0154 or 411.9926).

LarC: The above LarE reaction mixture was incubated with LarC (50 μM), CTP (1 mM), $NiCl_2$ (0.1 mM), and CoA (1 mM) for 2 h at RT and the proteins were heat denatured (60 °C, 5 min) and removed using a 10kDa 1.5ml centrifugal filter. The filtrate was concentrated to 50 μl and analyzed by LC-MS to determine the formation of the (SCS)Ni(II) pincer product **16** ([M+1] = 467.9123).

Synthesis of Biotin-4-carboxybenzene sulfonyl azide (91): Biotin-4-carboxybenzene sulfonyl azide (**91**) was synthesized by Dr. Sameh Abdelwahed in our lab with the starting materials, 4-carboxybenzenesulfonazide and biotin hydrazide (Figure 5.15 and 5.16).

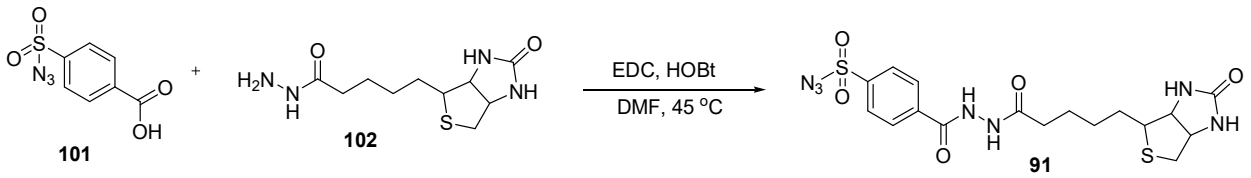


Figure 5.15: The synthetic route of Biotin-4-carboxybenzene sulfonyl azide (biotinSA, **91**).

4-Carboxybenzenesulfonazide (**101**, 104 mg, 0.46 mmol) was dissolved in DMF (20 mL), followed by addition of N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC, 96 mg, 0.50 mmol), 1-hydroxybenzotriazole (HOBt, 84 mg, 0.55 mmol), NEt_3 (0.12 ml, 0.86 mmol) and (+)-biotin hydrazide (**102**, 107 mg, 0.41 mmol). The reaction was stirred at 45 °C for 18 hr, then the solvent was removed in vacuo. The reaction mixture was subjected to flash chromatography (CH_2Cl_2:MeOH=19:1 to 3:2, v/v) and size exclusion chromatography using Sephadex LH-20 to afford **91** as a white solid (44% yield). 1H NMR (300 MHz, DMSO-D_6) δ 8.23 (d, *J* = 7.9 Hz, 2H), 8.11 (d, J =7.9 Hz, 2H), 4.51-4.65 (m, 2H), 3.28-3.38 (m, 1H), 2.78-3.22 (m, 2H), 2.16-2.24 (m, 2H), 1.51-1.69 (m, 4H), 1.22-1.33 (m, 2H); 13C NMR (300 MHz, DMSO-D_6) δ 173.9, 168.2, 164.8, 140.0, 133.3, 121.2, 119.5, 61.8, 60.3, 55.5, 39.7, 33.2, 28.2, 28.0, 25.0; ESI-MS (m/z): 468.1 (M+1).

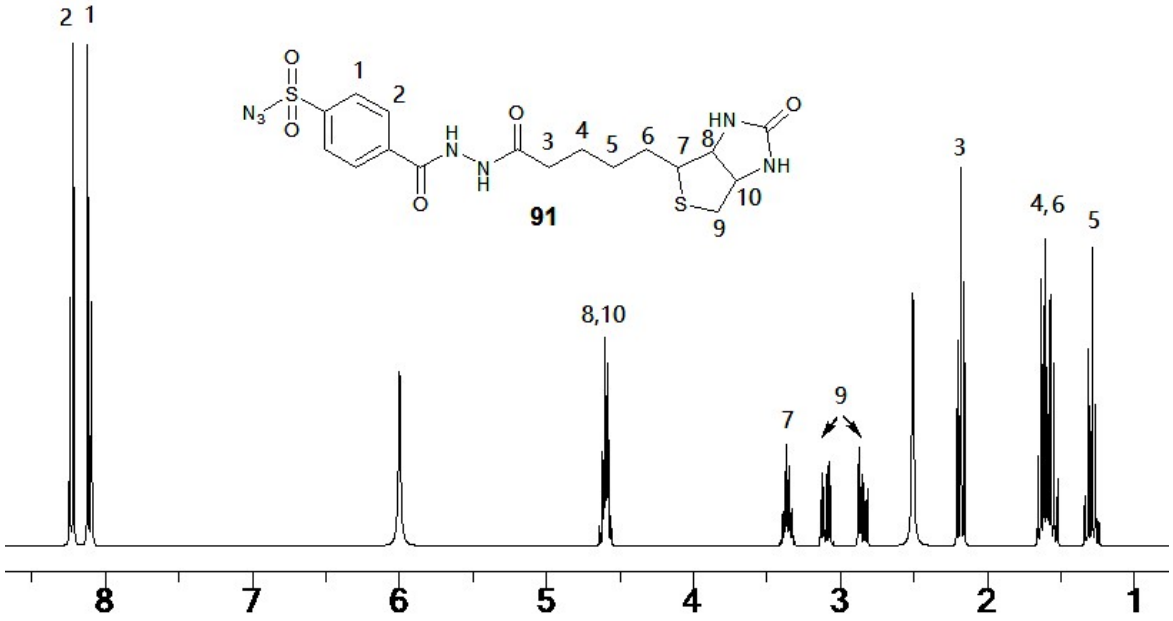


Figure 5.16: ^{1}H NMR (300 MHz, DMSO-d_6) of BiotinSA (**91**)

The synthesis and characterization of Lissamine Rhodamine B Sulfonyl Azide (LRSA, **90**) had been described in Chapter 4.

LRSA (90) labeling purified LarA-COSH (18) and LarE-COSH (14) and bacterial proteome: The LarA-COSH (**18**) was directly purified from the LarA(C-terminal StrepII-tagged)BCDE (plasmid pGIR112) overexpressed *Lc. lactis* cells following the protocol described above. The LarE-COSH (**14**) was not stable and it had to be freshly prepared in several steps before the labeling assay: 1) the LarE and LarB proteins were purified separately (plasmid pGIR026 only encoded the *larB* gene fused with a StrepII-tag at its C-terminal, and plasmid pGIR076 only encoded the C-terminal StrepII-tagged *larE* gene); 2) the LarE-COSH (**14**) was formed through the LarB/LarE coupled reaction using NaAD (**8**) as the substrate (50 μM LarE was incubated with 0.5 mM NaAD (**8**), 5 mM ATP, 20 mM $MgCl_2$, 10 mM $NaHCO_3$, 0.1 mM $NiCl_2$ and 20 μM LarB at RT for 30 min). The proteomes were freshly prepared, as mentioned

above, to maximally maintain the NAD derived thiocarboxylate cofactor covalently bound to LarA and LarE in the bacterial proteome.

The general LRSA (**90**) labeling procedure: 50 µl of ~200 µM freshly purified protein thiocarboxylates (LarA-COSH, **18**, or LarE-COSH, **14**), or the 200 µl bacterial proteomes from 200 ml cell cultures were solvent-exchanged into a PBS buffer (50 mM potassium phosphate, 300 mM NaCl, 6 M urea, pH 7.0). LRSA (**90**) was added (for purified protein samples, 3 equiv, ~2 µl, of 15 mM LRSA stock in DMSO; for proteome samples, 10 µl of the LRSA stock) and the sample mixtures were incubated in the dark for 15 min at RT. TCEP (4 µl of 250 mM TCEP stock (in water) for a purified protein sample; 6 µl of the stock for a proteome sample) was then added and the mixture was further incubated for 15 min in the dark at RT. The labeled samples were passed through a desalting column to reduce the free rhodamine background and loaded in a proper amount (eg. 20 µl) for the SDS-PAGE gel analysis. The rhodamine-labeled gel band could be observed with a Typhoon FLA 9500 instrument (Ex 530 nm/Em 580 nm).

To determine the LarA-COSH (**18**) labeling sensitivity with LRSA (**90)**, various concentrations (0.1 µM, 1 µM, 5 µl and 10 µM) of purified LarA-COSH (**18**) were added separately into the set amount of *Lc. Lactis* NZ3900 proteome (without the lar gene) for LRSA (**90**, 2 mM) labeling.

Biotin-4-carboxybenzene sulfonyl azide (91) labeling and enrichment of LarA-COSH (18) in larABCDE overexpressed *Lc. lactis* proteome (plasmid pGIR100): The proteome (200 µl, from 200 ml cell culture) was buffer-exchanged into the labeling PBS buffer (50 mM potassium phosphate, 300 mM NaCl, 9 M urea, pH 7.0). Then it was treated with 1 µl of **91** stock (250 mM in DMSO) and incubated at RT for 15 min. TCEP (6 µl) was added, and the mixture was incubated another 30 min followed by passing through a Biorad P6 size-exclusion

chromatography to remove the excess free biotin derived small molecules. A small portion of the labeled proteome was directly separated by an SDS gel and the biotinylated protein LarA-COSH (**18**) band was visualized by a Western blot assay (streptavidin-PE staining). The rest of the labeled protein mixture was loaded onto 300 µl streptavidin agarose resin slurry (Genscript) and incubated at RT for 30 min with shaking. The resin was then washed with 10 ml labeling PBS buffer. The enriched biotin-labeled LarA-COSH (**18**) could be eluted from the resin when incubating in the elution buffer (10 mM EDTA, 95% formamide, 5 mM free biotin) at 70 °C for 5 min. The eluted protein sample was visualized by running an SDS-PAGE gel (Coomassie blue staining). This LarA-COSH (**18**) enriched protein elute was also performed the trypsin in-solution digestion for peptide LC-MS/MS analysis to determine the identity of the enriched protein.

For the Western blot assay, the protein elute was first performed an SDS-PAGE gel separation and then the protein bands were transferred onto a PVDF membrane via a Biorad Trans-Blot Turbo System (100 V for 30 min). The membrane was washed with TBS buffer (2 mM Tris-HCl, 50 mM NaCl, pH 7.5), blocked with 3% BSA, washed with TBS buffer again and stained with the streptavidin-PE at 1:3000. This blot was then stored in TBS buffer and imaged using a Typhoon FLA 9500 instrument (Ex 546 nm/Em 580 nm).

For the trypsin in-solution digestion, the protein eluate was buffer exchanged into 50 mM ammonium bicarbonate buffer (pH 7.8), added 5mM DTT and incubated at 37°C for 1 h. Iodoacetamide was added to a final concentration of 15mM, and the mixture was incubated for additional 30 minutes in the dark at room temperature. Dilute the reaction with three volumes of 50mM Tris-HCl (pH 8) or 50mM ammonium bicarbonate (pH 7.8) to allow for trypsin digestion. The trypsin was added in 1:50 ratio and the mixture was incubated overnight at 37°C. The

digested peptide sample was injected for LC-MS/MS analysis. The data were analyzed by searching in the NCBI database using Bruker BioTools 3.2.

LarE-COSH (14) degradation time-course test using LRSA (90) labeling: A mixture of LarB (10 μM), NaAD (1 mM), $MgCl_2$ (4 mM), and $NaHCO_3$ (20 mM) in Tris·HCl buffer (100 mM; pH 8) was incubated at RT overnight to form the LarB product (**9**) stock solution. LarE (200 μM), ATP (10 mM), and $MgCl_2$ (20 mM) were mixed in 200 μl Tris·HCl buffer (100 mM; pH 8). 20 μl of the LarB product (**9**) stock solution was added to the LarE mixture. The LarE reaction mixture was removed in 50 μl portions at 15 min, 30 min and 2 h (after the LarB product (**9**) addition) for protein denaturing and LRSA (**90**) labeling.

HPLC parameters: The HPLC model was an Agilent 1200 or 1260 with a quaternary pump. The system included a diode array UV-Vis detector and eluted compounds were detected using absorbance at 254, 280, 260, 567, 220 and 340 nm. The analysis was performed on a ZORBAX Eclipse XDB-C18 column (4.6 x 150 mm, 5 μm particles, Agilent Technologies). Data were processed using ChemStation version B.04.01 SP1 (Agilent Technologies).

HPLC conditions:

A- Water

B- 10mM Ammonium acetate, pH 6.6

C- Methanol

HPLC method: (Flow rate = 1 ml/min)

0 min: 100% A; 7 min: 90% A, 10% B; 12 min: 25% A, 60% B, 15% C; 16 min: 25% A, 10% B, 65% C; 18 min: 100% B; 30 min: 100% B.

LC-MS parameters: LC-ESI-TOF-MS was performed using an Agilent 1260 HPLC system equipped with a binary pump and a 1200 series diode array detector followed by a

MicroTOF-Q II mass spectrometer (Bruker Daltonics) using an ESI source either in positive mode or negative mode. The analysis was performed on a Poroshell 120 EC-C18 column (3.0 x 100 mm, 2.7 μm particles, Agilent Technologies) for small molecules and a Luna® C5 100 Å, LC column (100 x 4.6 mm, 5 μm particles, Phenomenex Inc.) for proteins and a Phenomenex Synergi^{TM} LC column (50 x 2 mm, 2.5 μm particle size) for peptides.

LC conditions:

A: 5 mM ammonium acetate buffer, pH 6.6

B: 75% Methanol and 25% Water

LC method: (Flow rate = 0.4 ml/min, for both the positive and negative mode on MS)

0 min: 100% A, 0% B; 2 min: 100% A, 0% B; 12 min: 25% A, 75% B; 17 min: 25% A, 75% B; 18.5 min: 100% A, 0% B; 30 min: 100% A, 0% B.

Peptide LC conditions:

A: 0.1% formic acid in water

B: 0.1% formic acid in acetonitrile

LC method: (Flow rate = 0.4 ml/min, positive mode)

0 min: 95% A, 5% B; 5 min: 95% A, 5% B; 60 min: 55% A, 45% B; 90 min: 10% A, 90% B; 75 min: 10% A, 90% B; 90 min: 95% A, 5% B.

6. LABELING SMALL THIOCARBOXYLATES METABOLITES GENERATED DURING THE BIOSYNTHESIS OF THE NAD DERIVED PINCER COFACTOR

6.1 Metabolome and omics study

The metabolome is a term similar to "genome" and "proteome", which describes the whole set of metabolites (usually the small molecular intermediates and products generated during the metabolic process of life) existed in a biological sample, such as the cell extract from a specific bacterial strain. The dynamic of the metabolome is important in disease diagnosis. Even in ancient China, doctors already learned to figure out diabetes by observing whether the patient's urine could attract ants due to the high level of glucose.[90] To determine the function of an enzyme in a certain biosynthetic pathway, one can knockout the enzyme gene and check if there is some intermediate accumulation or shunt product generation. NMR is a useful way to characterize the structure of the target metabolite, though it may not be suitable for the trace amount metabolite detection in the metabolome. The high-resolution LC-MS method allows the isolation of the target metabolite from the metabolome complex and analysis of its mass information in high accuracy and sensitivity. Chemical modifications can facilitate the LC-MS detection by introducing a chromophore, improving the ionization efficiency, or generating more informative ms/ms fragments. In this chapter, two small thiocarboxylated metabolites (**12** and **15**, see Figure 6.1, in red boxes) in the LarA biosynthesis pathway were chemically modified by Lissamine Rhodamine B Sulfonyl Azide (LRSA, **90**) and dansyl azide (**92**) (Figure 6.2) and the labeling products were characterized by LC-MS. It has been the first time using the sulfonyl azide reagents to label trace amount of small thiocarboxylates (in μM scale) in the bacterial

metabolomes with the existence of much higher concentration (in mM scale) of many other metabolites.

Powerful computational techniques enable the comprehensive analysis of biological problems based on their genomic, transcriptomic, proteomic, and metabolomic data and their interactions. Such "omics" approaches open up the door of unraveling the timely and spatial complexity of biosystems (e.g., the evolution of species and the gut microbial community).

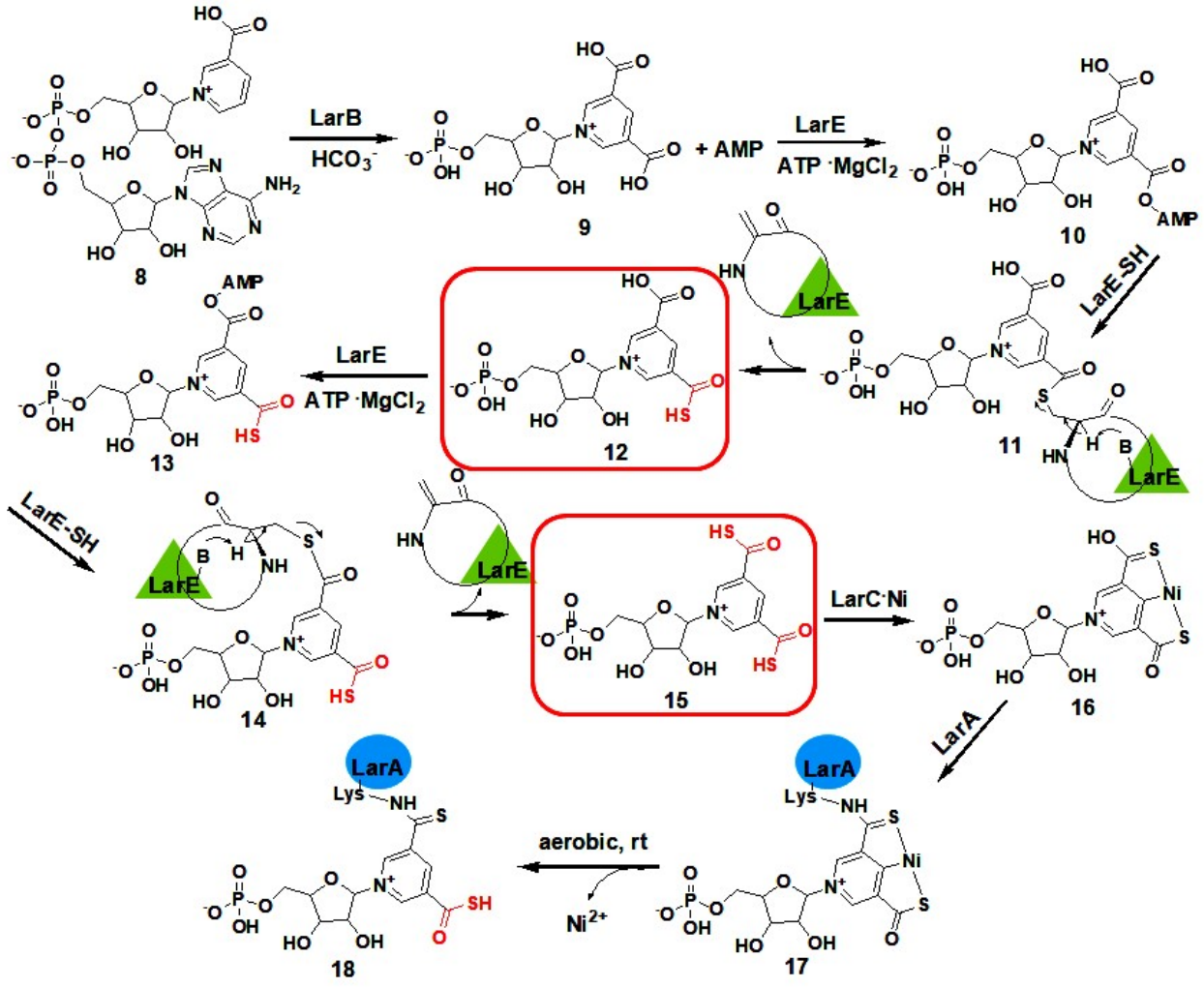


Figure 6.1: The two small thiocarboxylate metabolites in the red boxes, **12** and **15**, were selectively labeled in the LarA cofactor biosynthesis pathway.

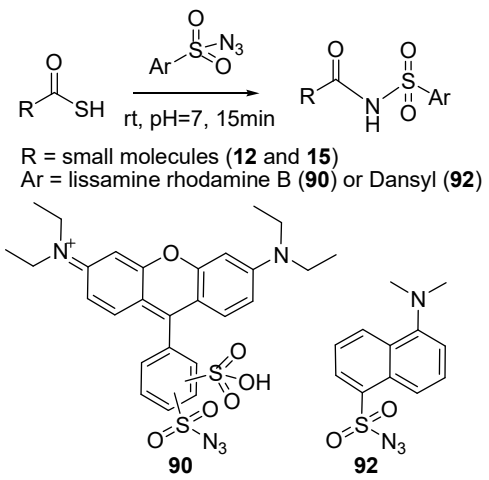
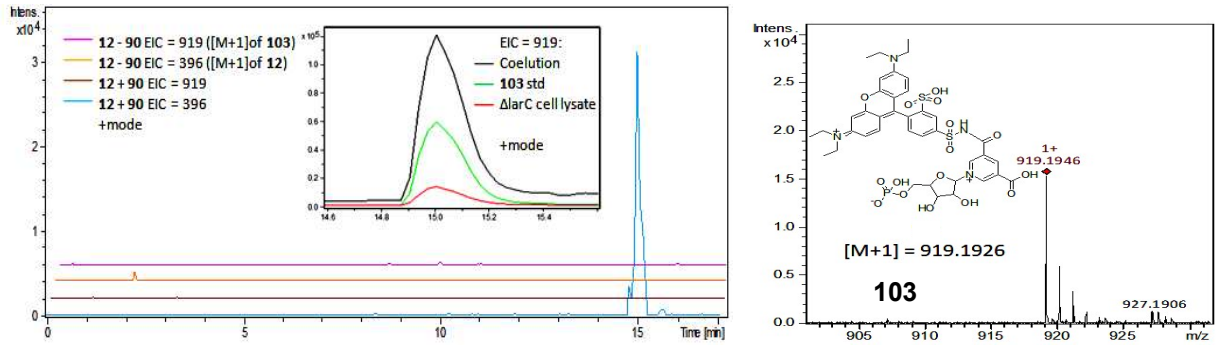


Figure 6.2: The key sulfonyl azide-thiocarboxylate reaction using Lissamine Rhodamine B Sulfonyl Azide (LRSA, **90**) and dansyl azide (**92**) to selectively label the two small thiocarboxylate metabolites (**12** and **15**) generated from LarA cofactor biosynthesis pathway (Figure 6.1, red-boxed).

6.2 Results and discussions

Besides the labeling of two protein thiocarboxylates **14** and **18** (see Figure 6.1, details were described in Chapter 5), the two small thiocarboxylate metabolites **12** and **15** were also produced during the LarA cofactor **16** biosynthesis. To confirm their interaction with LRSA (**90**, Figure 6.2), **12** and **15** were biosynthesized from LarB/LarE coupling reaction with NaAD (**8**) as reported[88]. The production of **12** or **15** could be controlled by adjusting the NaAD (**8**): LarE concentration ratio to 1:1 or 1:2.

A)



B)

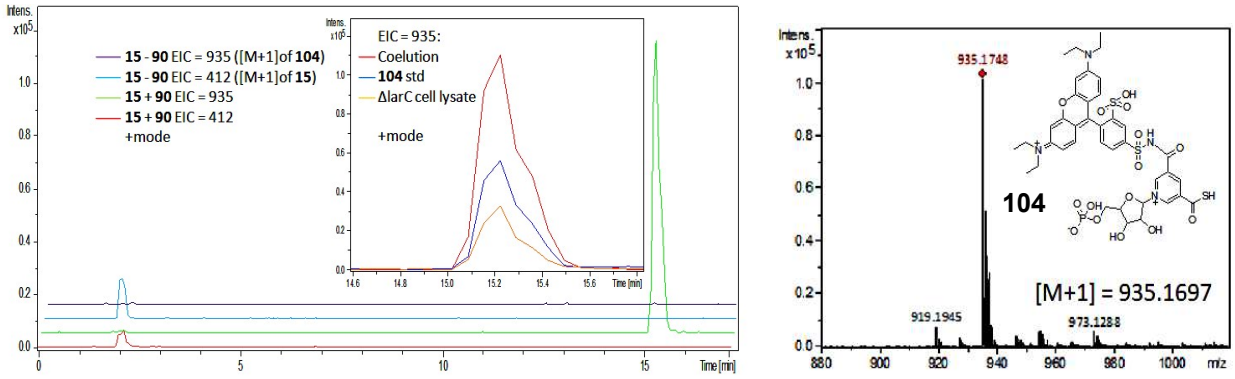


Figure 6.3: LC-MS analysis of LRSA (**90**) trapped small molecule thiocarboxylates **12** and **15** biosynthesized standards and from the bacterial cell lysate. A) Left: Extracted ion chromatography (EIC) of **12** and its rhodamine adduct **103** before and after **90** addition. Right: The structure and accurate mass of **103**. Inset: EIC of **103** detected in the LarAB(ΔC)DE overexpressed *Lc. lactis* cell lysate could coelute with **103** standard. B) Left: Extracted ion chromatography (EIC) of **15** and its rhodamine adduct **104** before and after **90** addition. Right: The structure and accurate mass of **104**. Inset: EIC of **104** detected in the LarAB(ΔC)DE overexpressed *Lc. lactis* cell lysate could coelute with **104** standard.

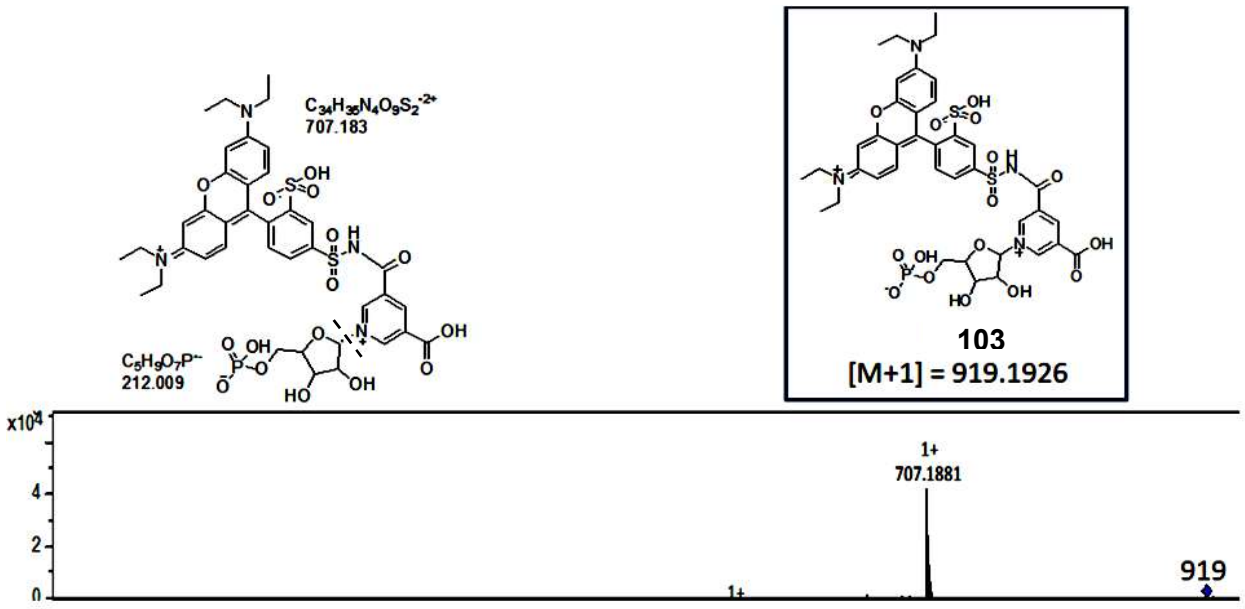


Figure 6.4: MS/MS fragmentation of **103** in positive mode

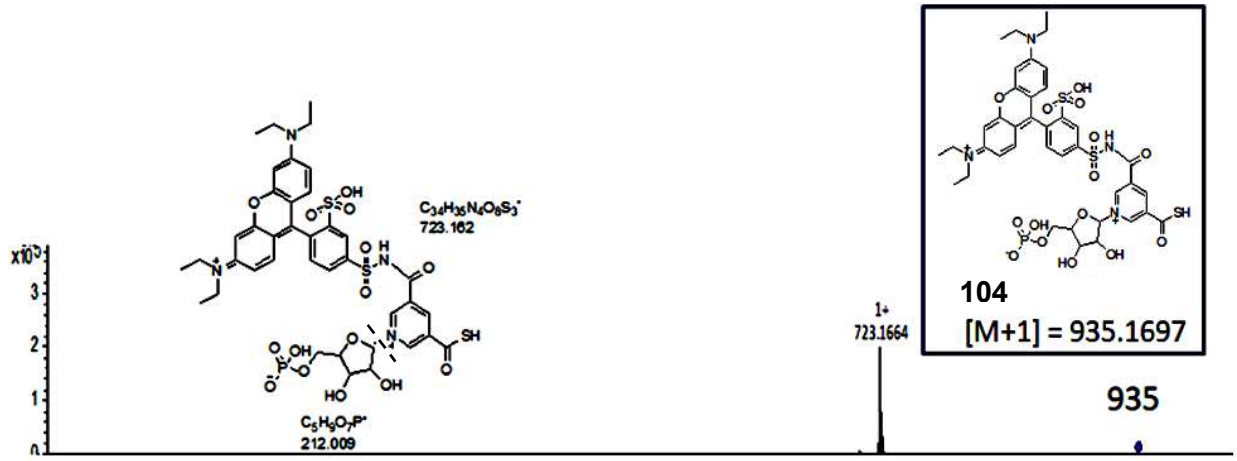


Figure 6.5: MS/MS fragmentation of **104** in positive mode

90 was then reacted with **12** and **15** standards and the products (**103** and **104**) were detected by LCMS. Both **103** and **104** extracted ion chromatogram (EIC) peaks appeared at the retention time of around 15 minutes with high mass accuracy (Figures 6.3A and 6.3B)and

reasonable ms/ms fragment assignments (Figures 6.4 and 6.5). The positive charge of the rhodamine group of **103** and **104** improved the ionization and the m/z intensity of **103** and **104** became much higher than **12** and **15** at the ESI positive mode, which increased the indirect detection limit of **12** and **15**. Only one of the two thiocarboxylate groups of **15** was labeled when 2 eqiv. LRSA (**90**) was added**,** due to the competitive side reaction of LRSA hydrolyzation. Another fluorescent reagent, Dansyl azide (**92**), was hydrolyzed much slower than LRSA (**90**) and EIC peaks of both mono- and di- adducts between **15** and **92** (**105** and **106**) were obtained with high accuracy (Figure 6.6). The ms/ms analysis also matched the structures of **105** and **106** very well (Figures 6.7 and 6.8).

LRSA (**90**) was further used to detect the **12** and **15** released in the cell lysate of pGIR100 (LarABCDE), pGIR200 (LarΔABCDE), and pGIR500 (LarABΔCDE) overexpressed *Lc. lactis* as well as the wild-type *L. plantarum*. Only pGIR500 (ΔlarC) overexpressed *Lc. lactis* cell lysate showed clear **103** and **104** EIC mass peaks that could coelute with the standards (the insets of Figure 6.3A and 6.3B). LarC inserted a nickel to **15** to produce **16**. The nickel ion made the thiocarboxylate group of **16** vulnerable to oxidative degradation and could not be labeled efficiently by **90** anymore. Without LarC in the proteome, LarB and LarE together led to the accumulation of **12** and **15** in the cell lysate. These two small thiocarboxylate compounds were much more oxidation resistant than **16** and thus could be successfully detected by **90**. When **16** was covalently bound inside the LarA enzyme, the protein structure may have helped prevent the cofactor from further oxidation process. Therefore, the oxidation of **17** could stop at the Ni losing step with the formation of a relatively stable protein thiocarboxylate, **18**, for **90** labeling.

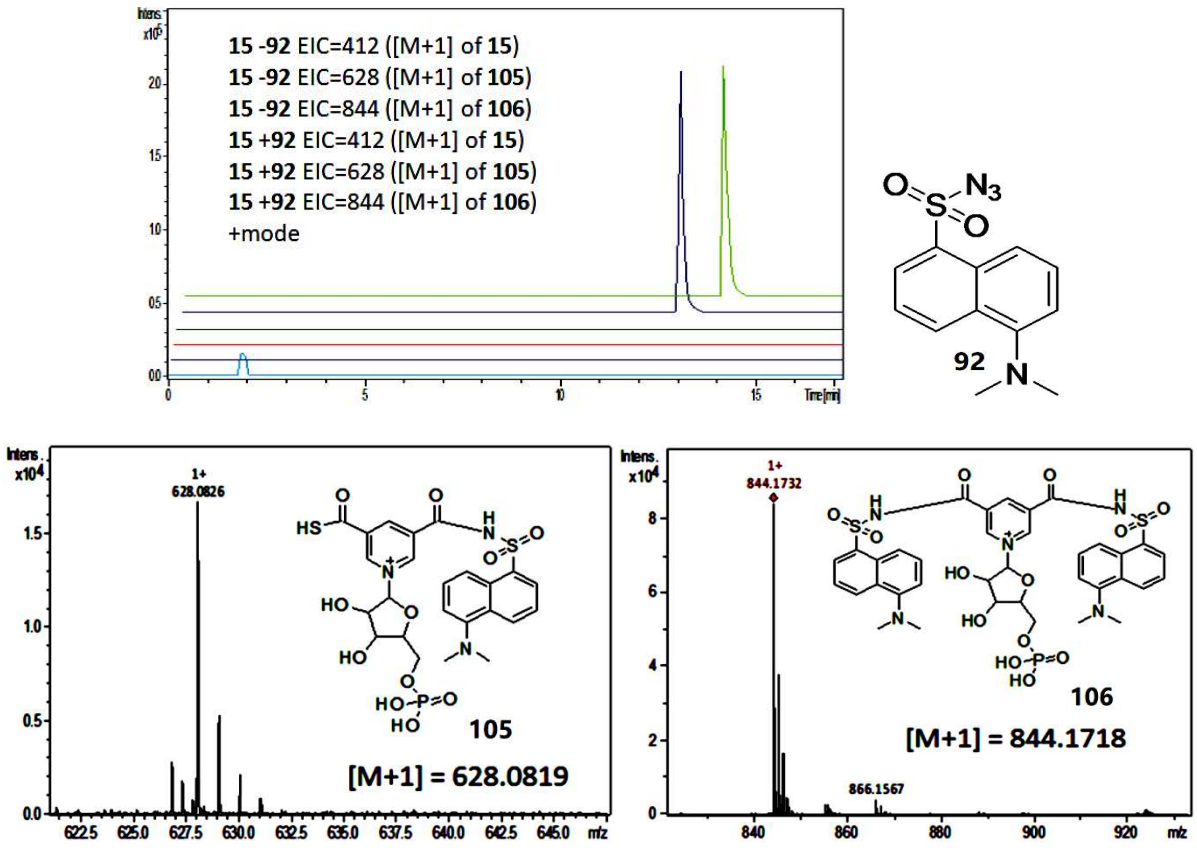


Figure 6.6: LC-MS analysis of Dansyl azide (**92**) labeled the small di-thiocarboxylate compound **15** and formed mono- and di- dansyl adducts **105** and **106**.

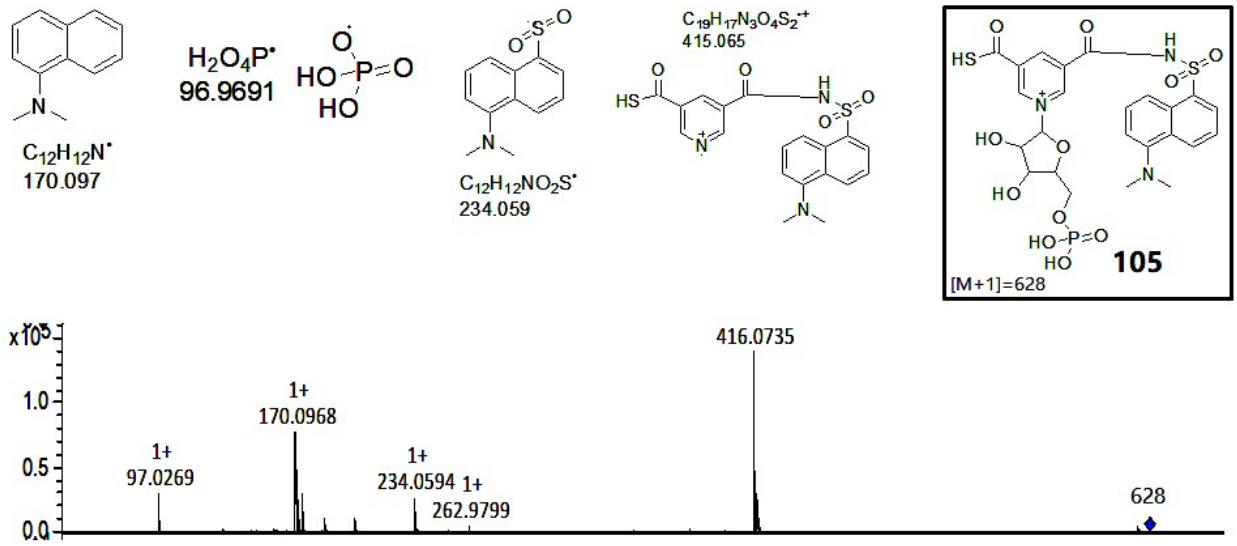


Figure 6.7: MS/MS fragmentation of **105** in positive mode.

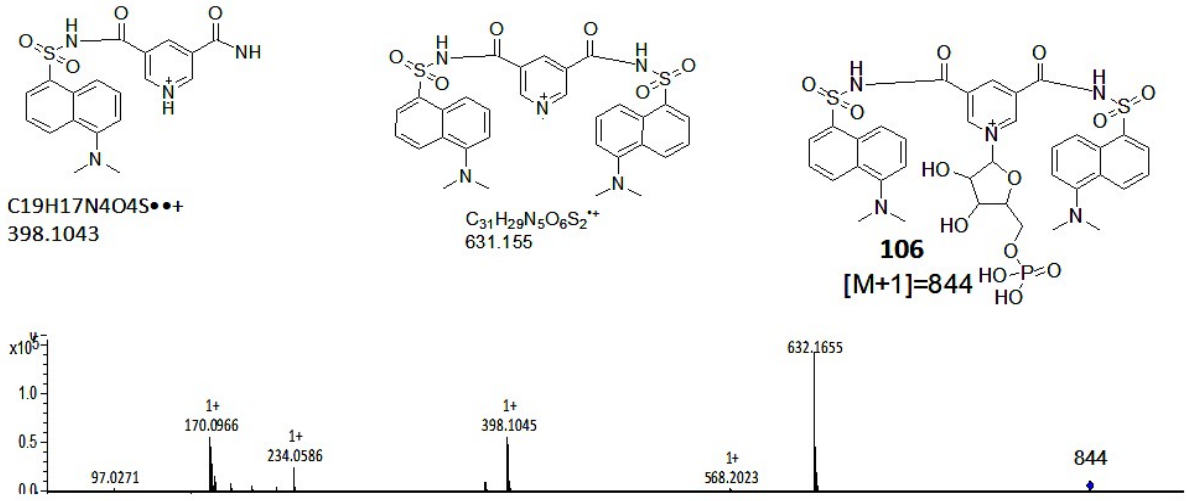


Figure 6.8: MS/MS fragmentation of **106** in positive mode.

6.3 Conclusion

The two small thiocarboxylate metabolites (**12** and **15**) involved in the NAD derived cofactor NiP2TMN (**16**) biosynthesis pathway were chemically labeled by LRSA (**90**) and Dansyl azide (**92**) and detected in the metabolome of LarAB(ΔC)DE overexpressed *Lc. lactis*

strain by LC-MS, LC-MS/MS analysis. The labeling product (**103-106**) peaks were coeluted with their standards, and their MS and MS/MS information also matched very well. It is the first time using the fluorescent sulfonyl azides to selectively label the extremely low concentration small thiocarboxylate metabolites in bacterial metabolomes. The labeling products (**103-106**) were ionized much better than the non-labeled small thiocarboxylate cofactors (**12** and **15**), and the mass intensity of **103-106** were ~10 times higher than **12** and **15**. The labeling method had improved the sensitivity of the small LarA thiocarboxylate cofactor detection. Combined with the protein thiocarboxylates labeling results, the sulfonyl azides reagents may have general applications in detecting new pathways using LarA cofactor but with different functions.

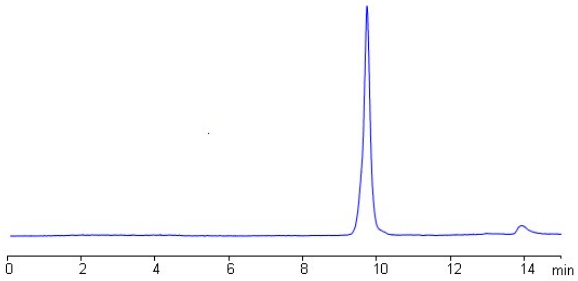
6.4 Materials and methods

Materials: All chemicals were purchased from Sigma-Aldrich unless otherwise mentioned. LB-Lennox broth was from EMD Millipore. Kanamycin and IPTG were obtained from Lab Scientific Inc. Chloramphenicol was from Fisher Scientific. QIAprep Spin Miniprep Kit was obtained from QIAGEN. Gravity flow Strep-Tactin® Superflow® high capacity columns were purchased from IBA GmbH. D-chloroform, D6-DMSO, D4-methanol were purchased from Cambridge Isotope Laboratories. 2.5L baffled ultra yield flask for overexpressing proteins were from Thomson Instrument Company. The plasmids and strains used in this project were kindly provided from Dr. Robert P. Hausinger's lab (Table 5.1).

Biosynthesis of the standards of two small thiocarboxylates (12 and 15): 12 and **15** standards biosynthesis and purification (the protocol was kindly provided by Dr. Benoît Desguin, a former post-doc researcher in Prof. Hausinger's lab). The reaction mixture contained LarB (50 μM), LarE (300 μM), NaAD (**8**) (For making **12**, 500 μM; for making **15**, 130 μM), ATP (5 mM),

$MgCl_2$ (20 mM), and $NaHCO_3$ (10 mM) in Tris·HCl buffer (200 mM; pH 7) in a final volume of 6 ml. It was incubated overnight at RT. The proteins in the mixture were precipitated by heating at 60 °C for 5 min. The crude product solution was passed through a 1.5ml 10K NMWL centrifugal filter and dried by lyophilization. The crude product was redissolved in 2 ml methanol to remove the salts (insoluble in methanol and removed by centrifugation). The methanol was evaporated using a Savant™ SpeedVac (Thermo Fisher) and the crude product was redissolved in 1 ml water and purified through an anion exchange chromatography using an AKTA system (GE Healthcare Life Sciences). A 5 ml HiScreen Q HP column from GE Healthcare was used with the flow rate of 1 mL/min. The wash buffer was Tris-HCl buffer (pH8; 20 mM), and the 1 M NaCl in the wash buffer was used for elution. The desalted crude product aqueous solution (1 ml) was loaded on the column and washed with 2 column volumes of wash buffer. Then the product was isolated and eluted by a gradient of 1 M NaCl in 80 min. Both **12** and **15** were monitored by their UV absorbance at 330 nm and eluted in ~350 mM NaCl similarly. The fractions were collected and the **12** and **15** (with the same HPLC retention time) were distinguished by MS (Figure 6.9).

A)



B)

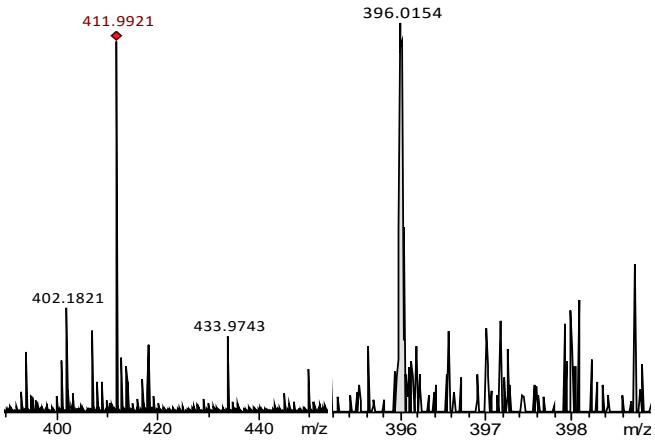


Figure 6.9: LC-MS analysis of purified **12** and **15** standards. A) Both **12** and **15** had the same retention time (~10 min) in HPLC analysis; B) **12** and **15** were differentiated by their mass. (**15:** [M+1] = 411.99; **12:** [M+1] = 396.02)

Bacterial metabolome preparation: Bacterial cell extracts were obtained following the procedure mentioned in Chapter 5. The extracts were then heated at 60 °C for 10 min to denature the Lar proteins for the efficient releasing of small thiocarboxylate metabolites (**12** and **15**). The proteomes also precipitated and were removed by centrifugation. The supernatant was passed through a 1.5ml 10K NMWL centrifugal filter unit to remove the soluble proteins.

Synthesis of Dansyl Azide (92): Dansyl chloride (**107**, 25 mg, 92 µmol, Figure 6.10) was dissolved in 10 mL of acetone in a round-bottomed flask wrapped in aluminum foil. Sodium azide (29 mg, 446 µmol, 5 equiv.) was added, and the reaction mixture was stirred at RT for 24

h. The solvent was removed in vacuo, and the residue was redissolved in dichloromethane. The resulting solution was washed with water, dried over anhydrous $MgSO_4$, and filtered, and the solvent was removed in vacuo to give Dansyl azide (**92**, 90% yield). A 15 mM stock solution in dimethylsulfoxide (DMSO) was prepared and stored in 1 ml aliquots at -20 °C in the dark. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.80 (d, 1H), 8.51 (d, 1H), 8.38 (d, 1H), 7.73 (t, 1H), 7.62 (t, 1H), 7.33 (d, 1H), 2.88 (s, 6H); 13C NMR (300 MHz, DMSO-D_6) δ 154.1, 134.0, 132.3, 129.3, 126.4, 124.5, 123.9, 121.3, 120.7, 115.8, 40.5; ESI-MS (m/z): 277.1 (M+1).

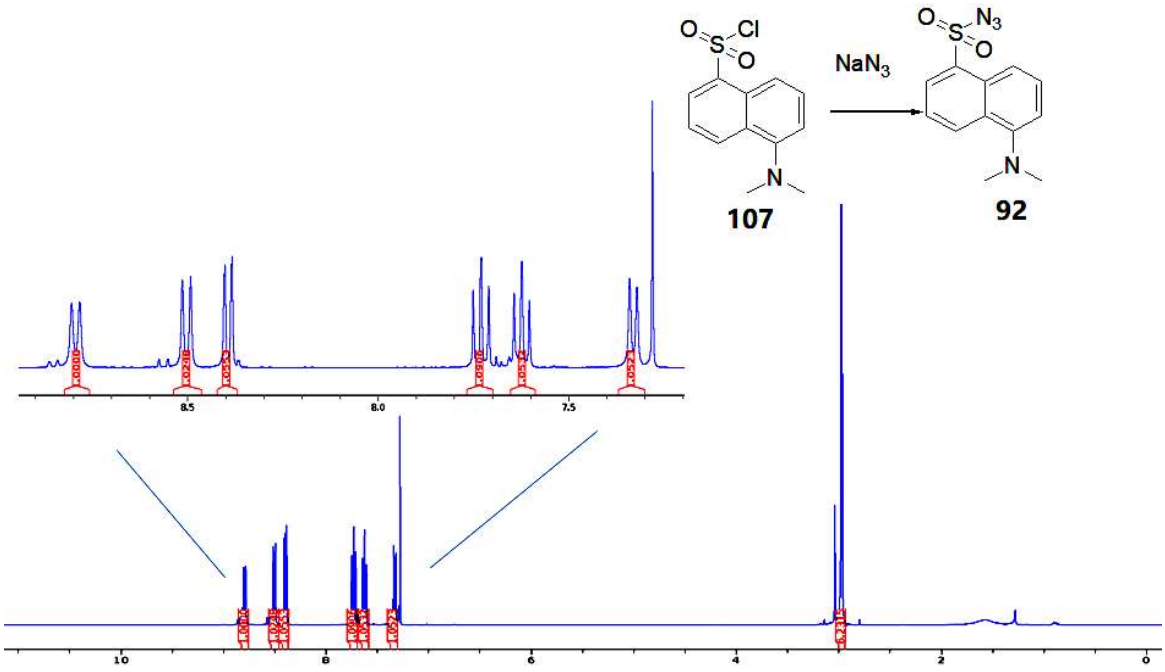


Figure 6.10: ^{1}HNMR of Dansyl azide (**92**).

Labeling conditions for 12 and 15 stands and in bacterial metabolomes: For purified standards labeling, **12** or **15** was dissolved in water or PBS buffer (~100 µM). A labeling reagent (LRSA, **90** or DanA, **92**) was added to the solution to make a final concentration 400 µM.

The labeling mixture was incubated in the dark for 15 min at RT. TCEP (1 mM final concentration) was added, and the mixture was stored in the dark for another 30 min. The sample was analyzed by LC-MS and LC-MS/MS using the positive mode.

For the bacterial metabolome labeling, the metabolome solution was prepared as discribed above and passed through a C18-preparative HPLC column in a 30 min method to remove the salts and undergo a rough isolation/enrichment of the target metabolites. Fractions between 9-12 min retention time were collected and combined for the labeling assays. In this way, **12** and **15** in the metabolome of 1 L cell culture were enriched and labeled by **90** (1 mM final concentration) for 15 min in the dark at RT, followed by the addition of TCEP (3 mM final concentration) and a further 30 min dark incubation at RT. The sample was analyzed by LC-MS and LC-MS/MS in the positive mode condition.

HPLC parameters: The HPLC model was an Agilent 1200 or 1260 with a quaternary pump. The system included a diode array UV-Vis detector and eluted compounds were detected using absorbance at 254, 280, 260, 567, 220 and 340 nm. The analysis was performed on a ZORBAX Eclipse XDB-C18 column (4.6 x 150 mm, 5 μm particles, Agilent Technologies) and a Supelco LC-18-DB (250 X 10 mm, 5 μm ID) column. Data were processed using ChemStation version B.04.01 SP1 (Agilent Technologies).

HPLC conditions:

A- Water

B- 10mM Ammonium acetate, pH 6.6

C- Methanol

HPLC method: (Flow rate = 1 ml/min)

0 min: 100% A; 7 min: 90% A, 10% B; 12 min: 25% A, 60% B, 15% C; 16 min: 25% A, 10% B, 65% C; 18 min: 100% B; 30 min: 100% B.

LC-MS parameters: LC-ESI-TOF-MS was performed using an Agilent 1260 HPLC system equipped with a binary pump and a 1200 series diode array detector followed by a MicroTOF-Q II mass spectrometer (Bruker Daltonics) using an ESI source either in positive mode or negative mode. The analysis was performed on a Poroshell 120 EC-C18 column (3.0 x 100 mm, 2.7 μm particles, Agilent Technologies).

LC conditions:

A: 5 mM ammonium acetate buffer, pH 6.6

B: 75% Methanol and 25% Water

LC method: (Flow rate = 0.4 ml/min, for both the positive and negative mode on MS)

0 min: 100% A, 0% B; 2 min: 100% A, 0% B; 12 min: 25% A, 75% B; 17 min: 25% A, 75% B; 18.5 min: 100% A, 0% B; 30 min: 100% A, 0% B.

7. BIOINFORMATICS STUDY ON POSSIBLE NEW THIS-COSH INVOLVED PATHWAYS

7.1 Bioinformatics study of ThiS/MoaD-COSH involved biosynthesis pathways

ThiS/MoaD-like protein thiocarboxylates (shortened as ThiS, ThiS-like, ThiS-COSH and MoaD-COSH, depending on the context) have been reported to transfer sulfur to various targets such as RNA nucleotides, amino acids, sugars, cofactors, antibiotics and other sulfur-containing natural products (Figure 7.1). These small ubiquitin-like sulfur carrier proteins (Figure 7.2) may participate in more intriguing sulfur new chemistry in vivo. By searching for their conserved genome neighbors that haven't been found in the known ThiS-COSH involved pathways (most commonly, the thiamin and molybdopterin biosynthesis) using powerful bioinformatics tools, our chance in finding the new sulfur-trafficking pathways has been increased dramatically.

A former graduate student, Yiquan Liu, used SEED database to find a conserved PLP-dependent enzyme annotated as threonine synthase next to a ThiS-like protein in many species without other thiamine/molybdopterin biosynthesis proteins around. She then reconstituted the protein activity and found that it actually functioned as a homocysteine synthase (Figure 7.3). In this chapter, I describe the bioinformatic analysis of ThiS-like protein family through the recently updated sequence similarity networks (SSNs) and genome neighborhood networks (GNNs) (https://efi.igb.illinois.edu/). Several enzyme genes were found conservely clustered with a *thiS* gene in genomes over multiple genera (> 30) for the first time. New reactions and mechanisms were proposed associated with these ThiS-like protein thiocarboxylates.

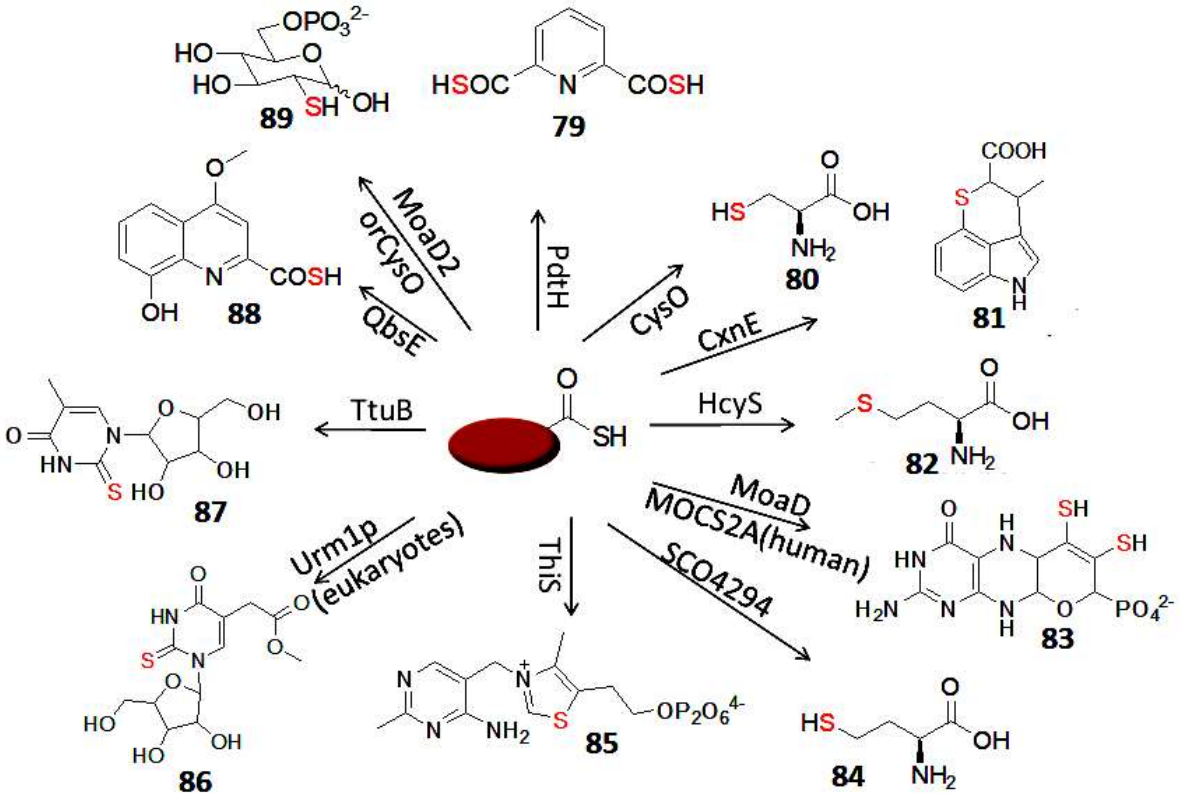


Figure 7.1: Summary of the reported ThiS-like sulfur carrier proteins involved in various natural product biosynthesis pathways. **79**: Pyridine dithiocarboxylic acid[21]; **80**: Cysteine[72]; **81**:Chuangxinmycin[73]; **82**: Methionine[26]; **83**: Molybopterin[74]; **84**: Homocysteine; **85**: Thiamin[75]; **86**: 5-methoxy-carbonyl-methyl-2-thiouridine[69]; **87**: 2-thioribothymidine[76]; **88**: Thioquinolobacterin[77]; **89**: 2-sulfur-glucose-6-phosphate[78].

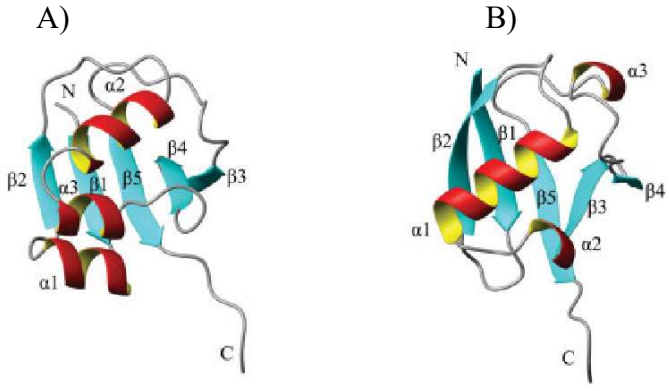
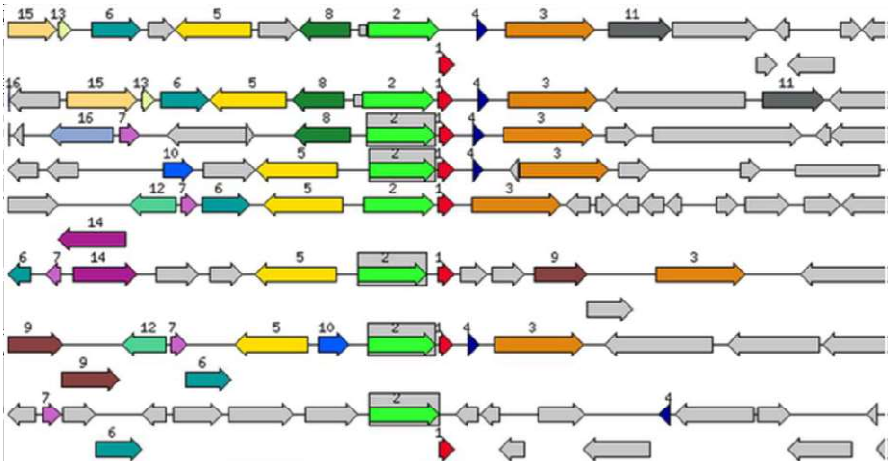


Figure 7.2: The structures of A) MoaD (a ThiS-like protein) from *Thermus thermophilus* (PDB: 1V8C) and B) ubiquitin from human (PDB: 1UBI).

A)



B)

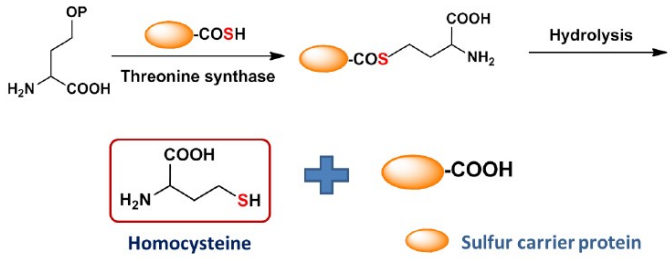


Figure 7.3: Bioinformatics assistance study of homocysteine synthase (used to be annotated as threonine synthase). A) Gene cluster of SCO4294 in *S. coelicolor* and analogs in other species. 1: ThiS/MoaD-family protein; 2: threonine synthase; 3: heat shock protein; 4: cold shock protein. B) The reaction. B) The annotated "threonine synthase" was actually a homocysteine synthase.

7.2 Results and discussion:

The sequence similarity network of the ThiS-like protein family (IPR003749: Sulfur carrier ThiS/MoaD-like InterPro family with 38482 proteins matched) was generated automatically using the EFI-EST online enzyme similarity tool (https://efi.igb.illinois.edu/efi-est/) based on an InterProScan sequence search (Figure 7.4) and visualized via Cytoscape software. Some reported ThiS-like proteins were color-labeled, showing that the sequence differences between ThiS-like proteins (ThiSs) were not always related to their function variations (maybe

due to their small protein size). The ThiSs with high sequence similarity (from the same node cluster, set the %sequence difference higher than 25% would be separated to a different node cluster) were found to transfer sulfur to different targets (e.g., CysO target: cysteine, QbsE target: thioquinolobacterin, and PdtH target: pyridine dithioarboxylic acid). While the ThiSs from different node clusters had the same sulfur acceptor (e.g., all MoaD/MoaD2s from different node clusters transferred sulfur to the molybdopterin cofactor). The discovery of possible new ThiS-involved pathways was more dependent on investigating their genome neighborhood. The genome neighborhood network (GNN) was generated automatically using the EFI-Genome Neighborhood Tool (EFI-GNT, https://efi.igb.illinois.edu/efi-gnt/). A typical GNN hub (ThiS-like protein)-spoke (Pfam neighbor) cluster was shown in Figure 7.5. ThiS candidates involved in new biosynthesis pathways should not contain thiamin or molybdopterin biosynthesis gene clusters (most abundant ThiS-involved pathways) around their gene neighborhood ($\pm$10 genes). The conserved protein neighbors of interest next to ThiSs should be well-distributed in multi-genera genomes (~30 genera or more in this study). I was able to discover several ThiS candidates satisfied with our screening rules and make hypotheses of their possible functions.

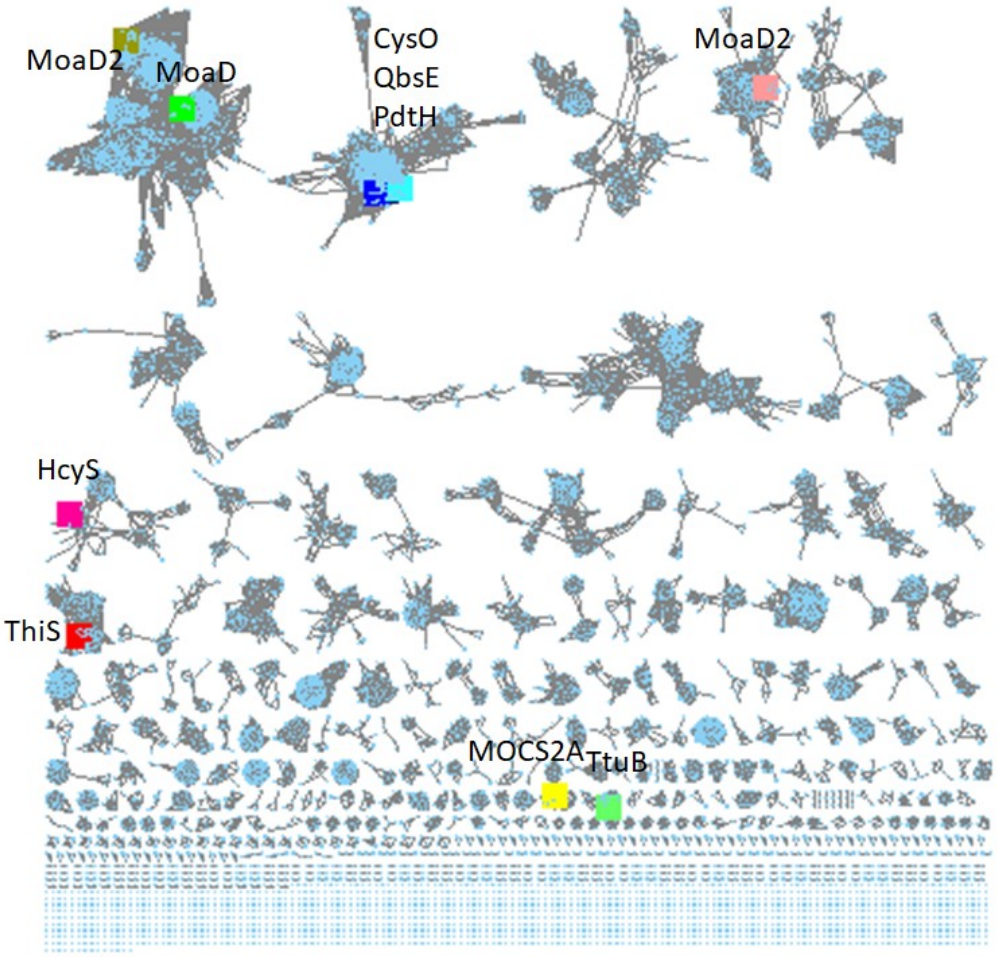


Figure 7.4: The sequence similarity network (SSN) of ThiS-like protein family (IPR003749: Sulfur carrier ThiS/MoaD-like, 38482 proteins matched). Some characterized ThiS-like proteins are labeled in colored squares.

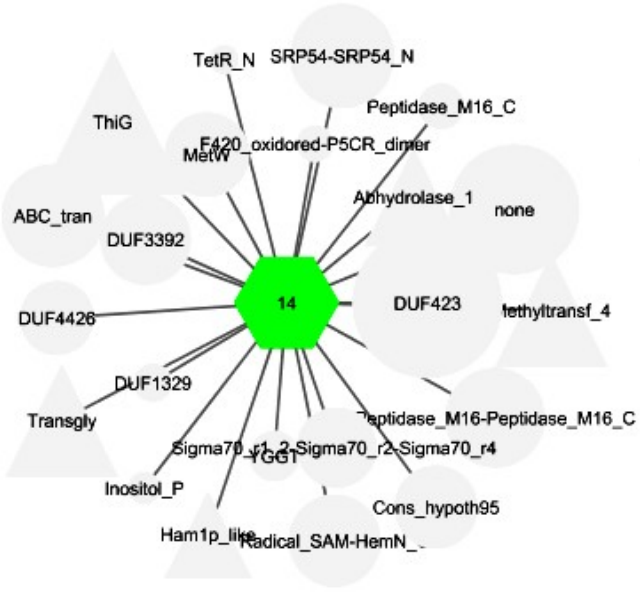


Figure 7.5: An example of hub (ThiS-like protein)-spoke (Pfam neighbor) cluster from the genome neighborhood network (GNN) of the ThiS-like protein family. There were no other thiamine or molybdopterin biosynthesis genes clustered with the ThiS-like protein except a thiazole synthase (ThiG). ThiG protein family was reported to catalyze reactions different from thiazole synthesis.[73,91]

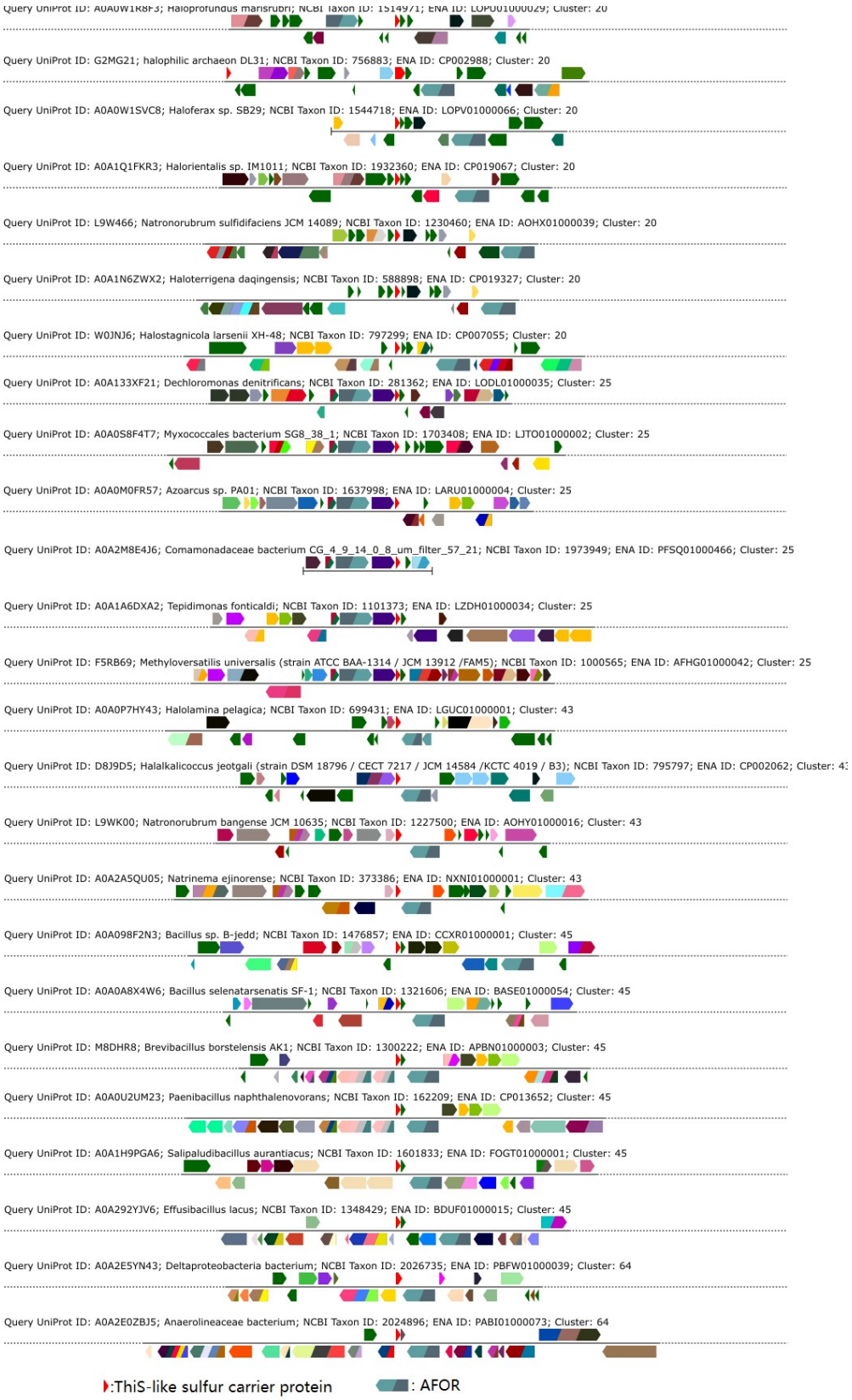


Figure 7.6: some examples of AFOR genes clustered with a ThiS-like protein gene in various genera genomes. (Ligands of conserved genes were shown at the bottom)

7.2.1 Aldehyde ferredoxin oxidoreductase (AFOR)

AFORs are widely distributed in many bacteria and archaea genera. They catalyze various aldehyde substrates to their corresponding carboxylate products.[92] The GNN analysis showed that around 1/3 of AFOR protein family members (more than 500 genomes, >>30 bacterial genera and more archaea genera) have a ThiS-like protein neighbor within $\pm$10 genes (see examples in Figure 7.6). In addition, most bacterial AFORs had a conserved FAD-dependent oxidoreductase adjacent to AFOR, annotated as "Pyridine nucleotide-disulfide oxidoreductase". While archaeal AFORs usually had a conserved radical SAM enzyme annotated as "tungsten-containing aldehyde ferredoxin oxidoreductase cofactor-modifying protein (Cmo)" (Figure 7.7 and 7.8). Generally, they don't co-exist in the same gene cluster with AFOR, which may indicate that their functions could be substituted from one to the other. The function of either the radical SAM or the FAD-dependent enzyme has not been figured out into details.

Figure 7.7: Examples of a conserved FAD-dependent oxidoreductase appeared next to the AFOR-ThiS gene clusters in over 30 bacteria genera.

Figure 7.8: Conserved Tungsten cofactor oxidoreductase radical SAM maturase

The predicted structure of a ThiS-like protein (UniProtKB: A0A0F7PGK1) next to AFOR was similar to a ubiquitin structure with a long C-terminal tail (Figure 7.9). AFORs had a tungsten-*bis*molybdopterin complex cofactor and a 4Fe-4S cluster at the active site. [93] The proposed mechanism used mono-oxo-W(VI) as the active form to catalyze the aldehyde oxidation (Figure 7.10), which agreed with one of the available crystal structures.[94] The activity of this enzyme has not been reconstituted. Some of the archaeal genome (e.g., *Pyrococcus furiosus*) has multiple AFORs[95] that may complicate the in vivo study and protein purification from the wild-type strain. Due to these limitations, the mechanism of AFOR catalyzed reaction has not been well explored.

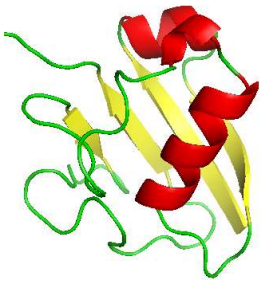


Figure 7.9: Predicted structure of a ThiS-like protein next to AFOR (UniProtKB: A0A0F7PGK1). The ubiquitin-like structure had a long C-terminal tail.

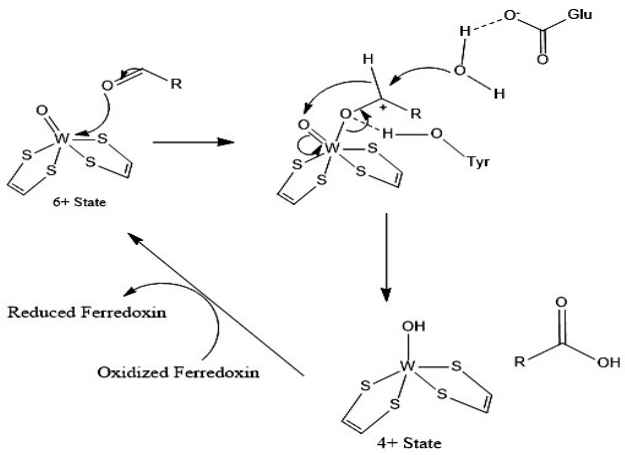


Figure 7.10: Proposed mechanism of the AFOR catalyzed reaction.

Tungsten (W, atomic number 74) and molybdenum (Mo, 42) are chemically analogous elements. The great similarity of their atomic properties may allow them to take the same catalytical strategies at the active site. Molybdenum-dependent enzymes have been widely investigated. Some of them also have a tungsten version with high structural similarity, such as the DMSO reductase family.[96] Many W/Mo-dependent enzymes require sulfur or selenium for their activity. Replacement of sulfur/selenium to oxygen inactivates the protein. The sulfur source for the oxo-sulfido-Mo(VI) cofactor of xanthine oxidase/dehydrogenase family was known as IscS (Figure 7.11A).[97] In addition, the oxo-sulfido-bis(dithiolene)tungsten(VI) Complexes have been synthesized using hydrosulfide (SH^-) as the sulfur donor (Figure 7.11B).[98] The sulfido- complexes were much more reative than the oxo-version of the same type complexes. Therefore, the active form of AFOR cofactor may also be possible to incorporate an active sulfido- group at the tungsten center for its activity reconstitution. The sulfur was not observed in the available crystal structures (PDB: 1B4N, 1B25, and 1AOR) maybe due to the fast hydrolysis of the sulfido- active form to the inactive oxo- form.[93,99] ThiS-COSH protein

thiocarboxylates are efficient sulfur donors in vivo with oxidation resistance property. It is highly possible that the conserved ThiS-like protein next to AFOR plays a crucial role as the sulfur donor for the active sulfido- cofactor generation.

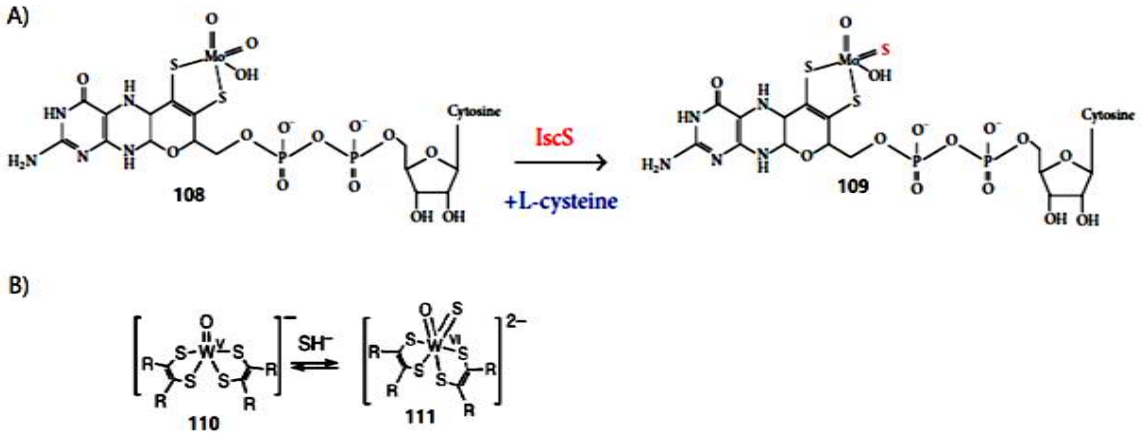


Figure 7.11: A) Sulfido incorporation step during the biosynthesis of oxo-sulfido Molybdenum Cofactor (Mo MPT) via IscS (the cofactor of xanthine oxidase/dehydrogenase family);[97] B) Synthesis of tungsten Oxo-sulfidobis(dithiolene)tungsten(VI) Complexes using hydrosulfide (SH^-).[98]

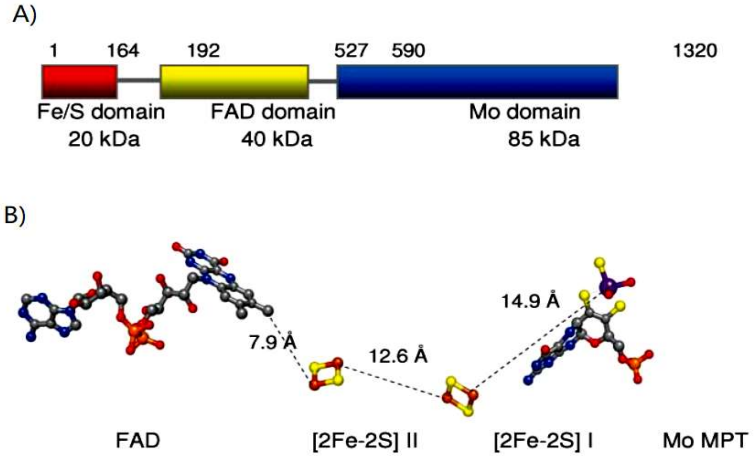


Figure 7.12: The FAD binding domain in the bovine milk xanthine oxidase. A) Schematic representation of the domain structure based on the primary sequence; B) Arrangement of the metal cofactors involved in electron transfer: Mo MPT, [2Fe-2S] I, [2Fe-2S] II and FAD.[96]

The earlier mentioned radical SAM maturase or the FAD-dependent oxidoreductase next to the AFORs may activate the AFOR enzymes by catalyzing the same sulfido-tungsten(VI) active complex formation in different organisms through either the radical mechanism or the hydride transfer mechanism (Figure 7.13A and B). The radical mechanism may generate an acylium ion. For the AFORs associated with FAD-dependent oxidoreductase, the two electrons grained from the aldehyde substrates need to be removed from the tungsten center for next round of reaction. They may transfer through the 4Fe-4S cluster, the FAD and finally to the NAD to form NADH and release from the active site (Figure 7.13C). Another oxo-sulfido-Mo(VI) cofactor dependent enzyme, xanthine oxidase (XO), has its own FAD binding domain as well as two 2Fe-2S clusters to fulfill similar electron transfer process more efficiently (Figure 7.12). For the radical SAM associated AFORS, the two electrons may travel from the tungsten center to the 4Fe-4S cluster of AFOR, then to the radical SAM 4Fe-4S cluster to initiate next round of Reductive cleavage of SAM to generate the 5'-adenosyl radical.[100] It may reflex the enzyme evolution trend from the AFOR with radical SAM maturase (majorly found in anaerobic archaea) to AFOR with FAD oxidoreductase (found in aerobic bacteria) to XO (from bovine milk).

On the other hand, there were AFORs with neither ThiS nor maturase around. It is possible that not all AFORs require sulfido-W(VI) cofactor for their activity. Or they may share the ThiS and maturase with other AFORs in the same genome. The sulfur source may also vary (IscS, or HS^-) in different organisms. The annotated "AFOR" may not be the real AFOR, but catalyze some reactions.

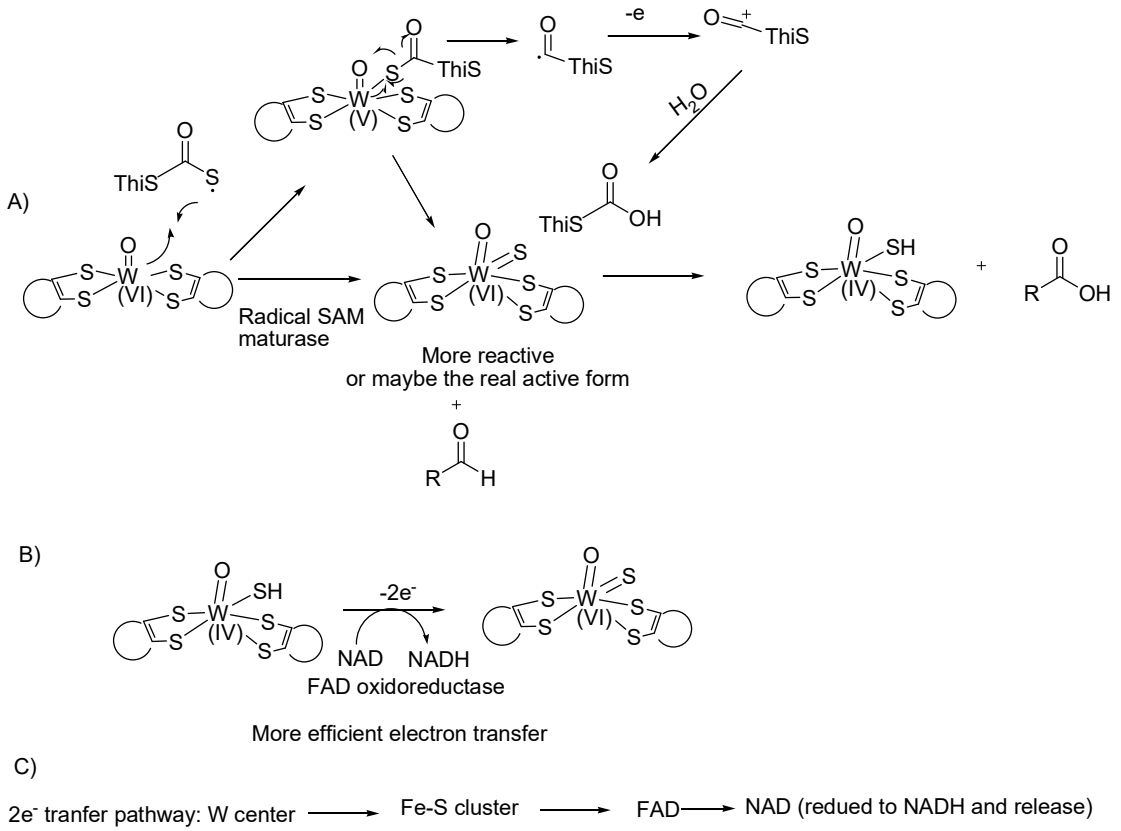


Figure 7.13: Proposed function of ThiS-like protein as a sulfur donor and incorporated to the tungsten cofactor though two types of maturases. A) Radical SAM maturase (acylium ion formation) B) FAD-dependent oxidoreductase that clustered to AFOR. C) Multi-step electron transfer for W(VI) formation catalyzed by FAD-dependent oxidoreductase.

7.2.2 tRNA (guanine-N(7)-)-methyltransferase (TrmB)

More than 100 genomes (>30 bacterial genera) were found with a ThiS-like protein next to TrmBs (examples see Figure 7.14). A highly conserved ThiG was also located next to ThiS. No other thiamin biosynthesis genes were found in this gene cluster.

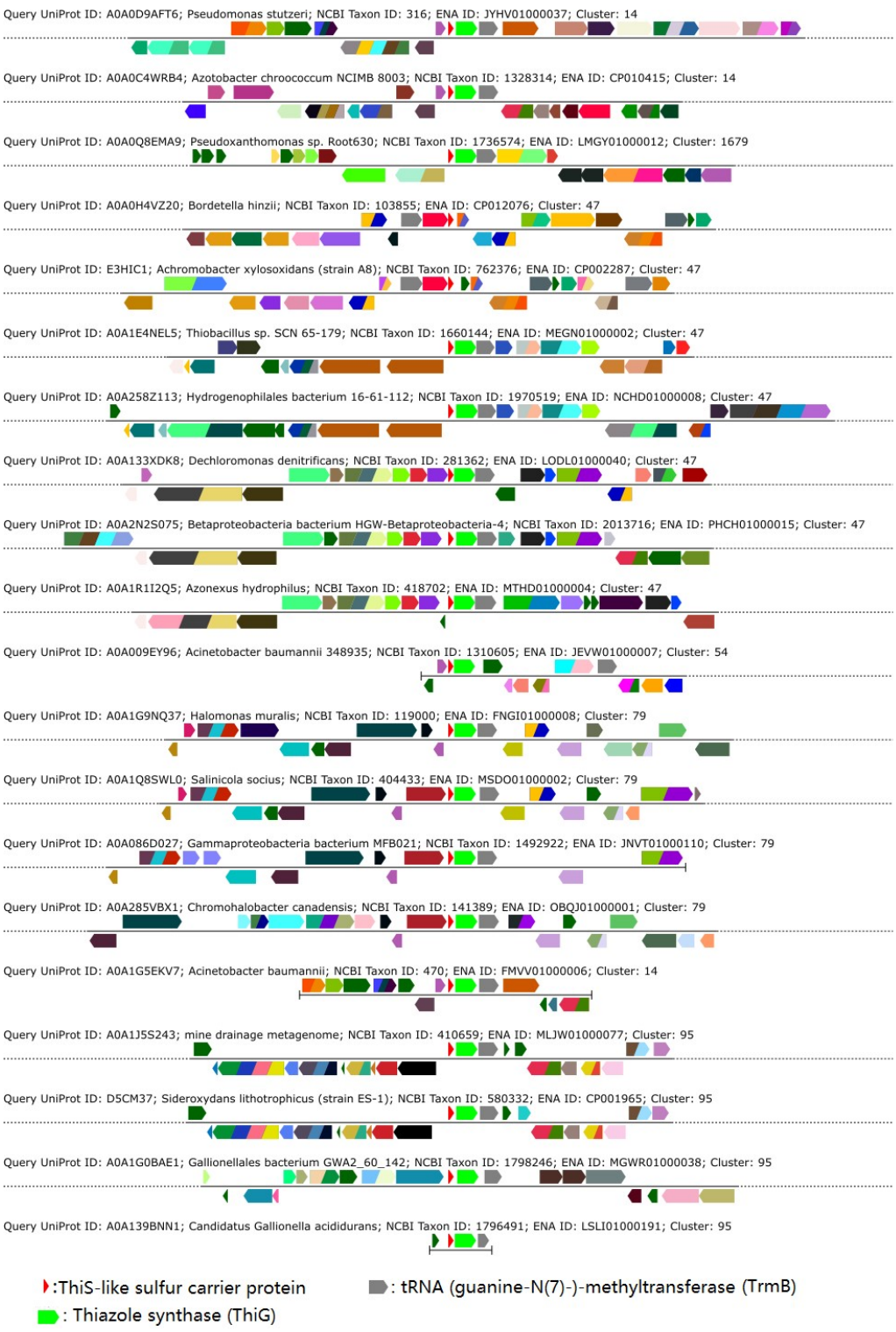


Figure 7.14: Examples of TrmB genes clustered with a ThiS-like protein gene in various genera genomes. (Ligands of conserved genes were shown at the bottom)

A predicted ThiS structure in this cluster showed a ubiquitin-like structure (Figure 7.15). The reaction TrmB catalyzed was shown in Figure 7.16A. The structure suggested a mechanism that the 7-N lone pair electrons directly attacked the methyl group of SAM since they were in perfect orientation.[101] Our proposed mechanisms used ThiS-COS$^-$ as a general base catalyst. The ThiG might also bind to the substrate to increase the ThiS-COS$^-$ attacking efficiency (Figure 7.16B, possibility 1 and 2). A third possibility was the reaction between ThiS-COS$^-$ and SAM (Figure 7.16B, possibility 3). After hydrolysis of the thioester, the product was methanethiol (**116**). Methanethiol is a molecule commonly generated by living organisms through a methioninase and is the substrate of a SAM-methyltransferase to form dimethylsulphide.[102,103] A SAM-dependent methyltransferase formed the same product using bisulfide (HS$^-$) as the substrate had been reported in plant.[102] Here we switched the substrate to ThiS-COSH. The initial found ThiG (thiamin thiazole synthase) catalyzed reaction was shown in Figure 7.16C)

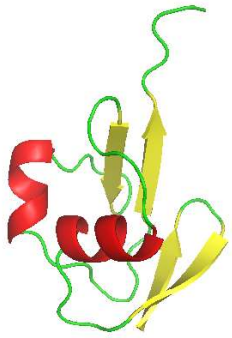


Figure 7.15: Predicted structure of a ThiS-like protein next to TrmB (UniProtKB: A0A0D9AFT6). It showed the ubiquitin-like structure with a long C-terminal tail.

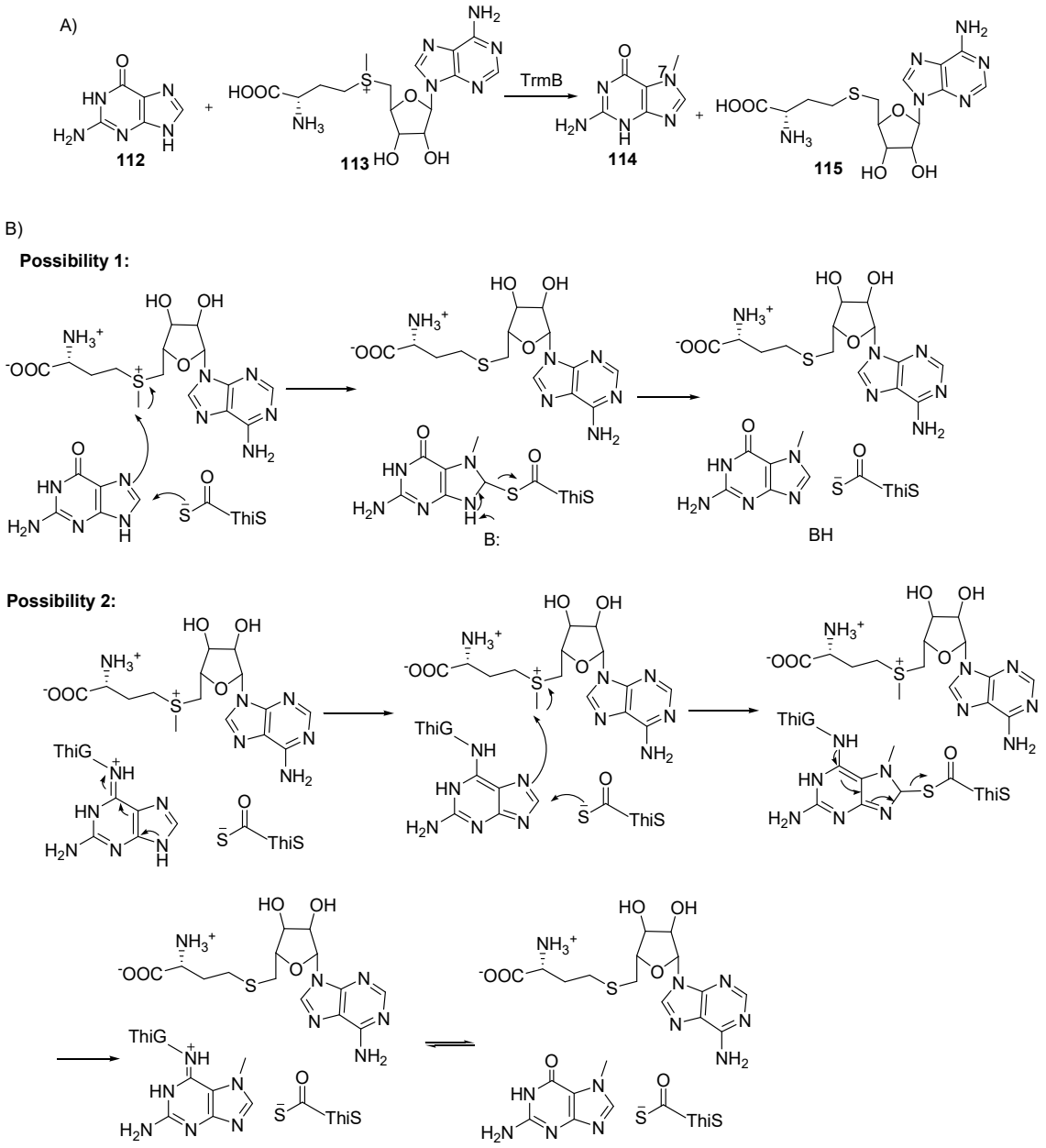


Figure 7.16: TrmB catalyzed reactions. A) The reported reaction catalyzed by TrmB without the contribution of ThiS-like protein. B) Three possible reactions that the conserved TrmB or ThiG may use ThiS-COSH as the general base or sulfur donor. C) Original thiamin thiazole synthase (ThiG) catalyzed reaction.

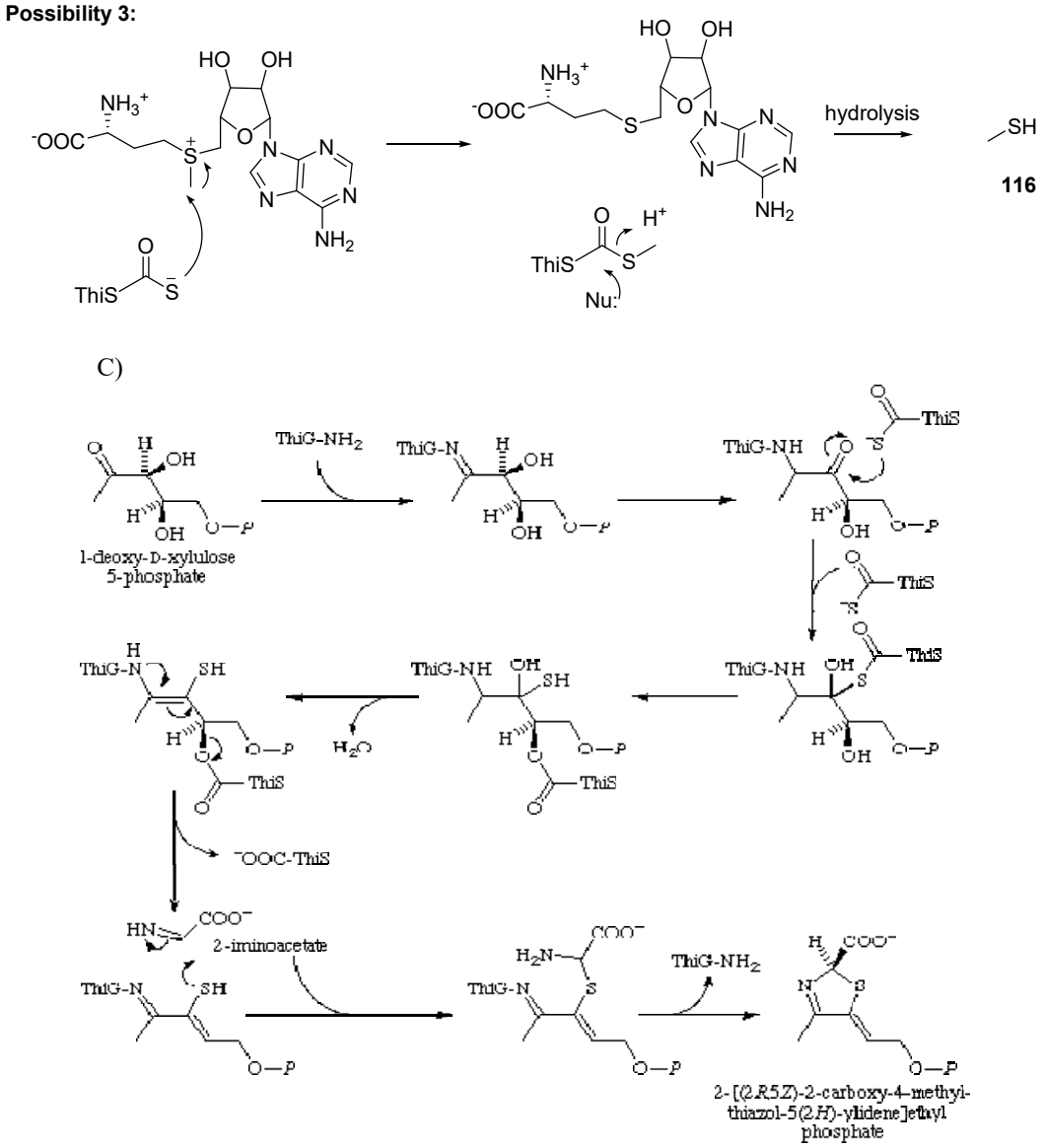


Figure 7.16: Continued

Figure 7.17: Examples of 3-dehydroquinate dehydratase genes clustered with a ThiS-like protein gene in various genera genomes. (Ligands of conserved genes were shown at the bottom)

7.2.3 3-dehydroquinate dehydratase (3QD)

More than 100 genomes (> 30 bacterial genera) had a ThiS-ThiG fusion protein next to 3QD (examples see Figure 7.17). Apart from 3QD, other highly conserved proteins in this gene cluster were: an Acetyl-CoA carboxylase biotin carboxylase subunit (AccC) and a Biotin carboxyl carrier protein of acetyl-CoA carboxylase (AccB). The ThiS part of a ThiS-ThiG sequence was used to predict the structure (Figure 7.18). A ubiquitin-like structure was obtained with a long C-terminal tail. With ThiS-COSH and ThiG together, 3QD may generate either the original product, 3-dehydroshikimate (**118**), or some other sulfur-containing products (**119-122**) (Figure 7.19 for various mechanisms).

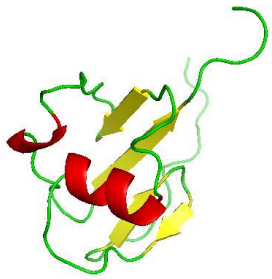


Figure 7.18: Predicted structure of the ThiS domain of a fusion protein ThiS-ThiG next to 3-dehydroquinate dehydratase (UniProtKB: L0DJ63). It showed the ubiquitin-like structure with a long C-terminal tail. The ThiS-ThiG fusion was common when clustered with 3-dehydroquinate dehydratase genes.

Thiocarboxylates had been used in peptide synthesis, and a radical-initiated dethiocarboxylation was observed (Figure 7.20A).[104] The carbonyl sulfide (COS) was dissolved in water to form thiocarbonate, that could potentially be catalyzed by biotin to form the corresponding thiocarboxylates. It may provide alternative ways to make thiocarboxylates other than activating a carboxylate group by ATP. This gene cluster did have a conserved biotin-

dependent carboxylase. The ThiS-COSH might be a possible thiocarbonate donor (Figure 7.20B). Carbonyl sulfide was also reported to mediate the prebiotic formation of peptides, which indicated its early cooccurrence with life.[105]

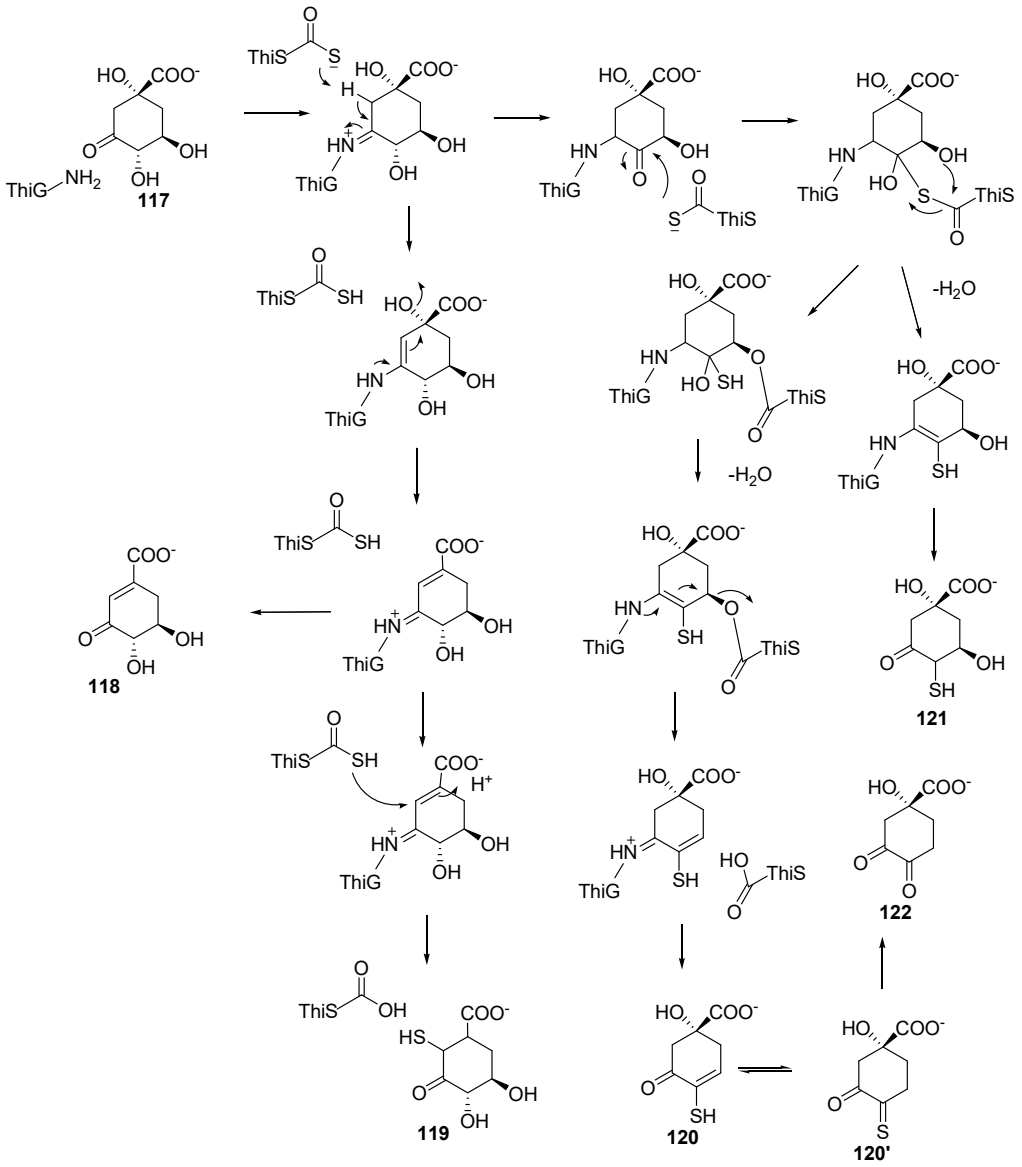
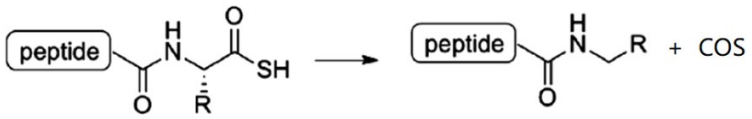


Figure 7.19: Proposed mechanisms and products of 3-dehydroquinate dehydratase reactions with the participation of a ThiS-ThiG protein.

A)



B)

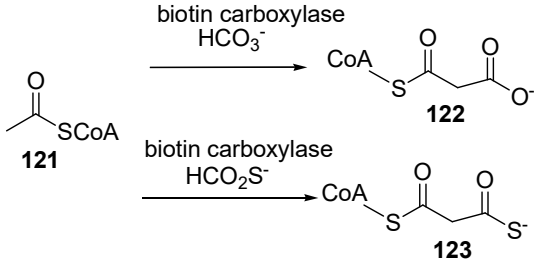


Figure 7.20: A) The reported dethiocarboxylation of peptide thiocarboxylates and CSO release.[104] B) Proposed formation of thiocarboxylates through a thiocarbonate addition catalyzed by a biotin carboxylase.

7.2.4 CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (CDP_OH_P)

CDP_OH_Ps were found adjacent to a ThiS/MoaD-like protein in more than 800 genomes (>>30 bacterial genera), often clustered together with a UvrABC system protein C and a moeE (examples were in Figure 7.21). No other thiamin or molybdopterin biosynthesis genes were found in this conserved gene cluster. The predicted structure of a ThiS/MoaD-like protein in this gene cluster showed the ubiquitin structural analogy (Figure 7.22).

Figure 7.21: Examples of CDP_OH_P genes clustered with a ThiS-like protein gene in various genera genomes. (Ligands of conserved genes were shown at the bottom)

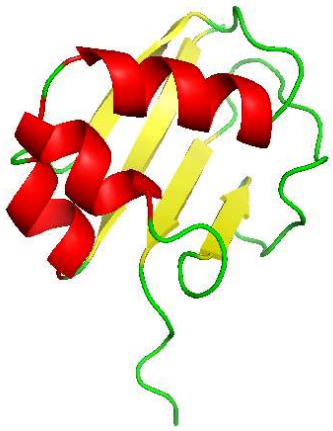


Figure 7.22: Predicted structure of a ThiS-like protein next to CDP_OH_P (UniProtKB: A0A159YZV6). It showed the ubiquitin-like structure with a long C-terminal tail.

The MoeE may catalyze the MoaD-COSH to cleave the phosphoester bond and make the reaction more efficiently (Figure 7.23B). Another conserved protein, UvrABC system protein C, was reported to cleave the damaged DNA at both 5' and 3' positions. MoeE/MoaD-COSH may make this cleavage more efficient (Figure 7.23C).

A)

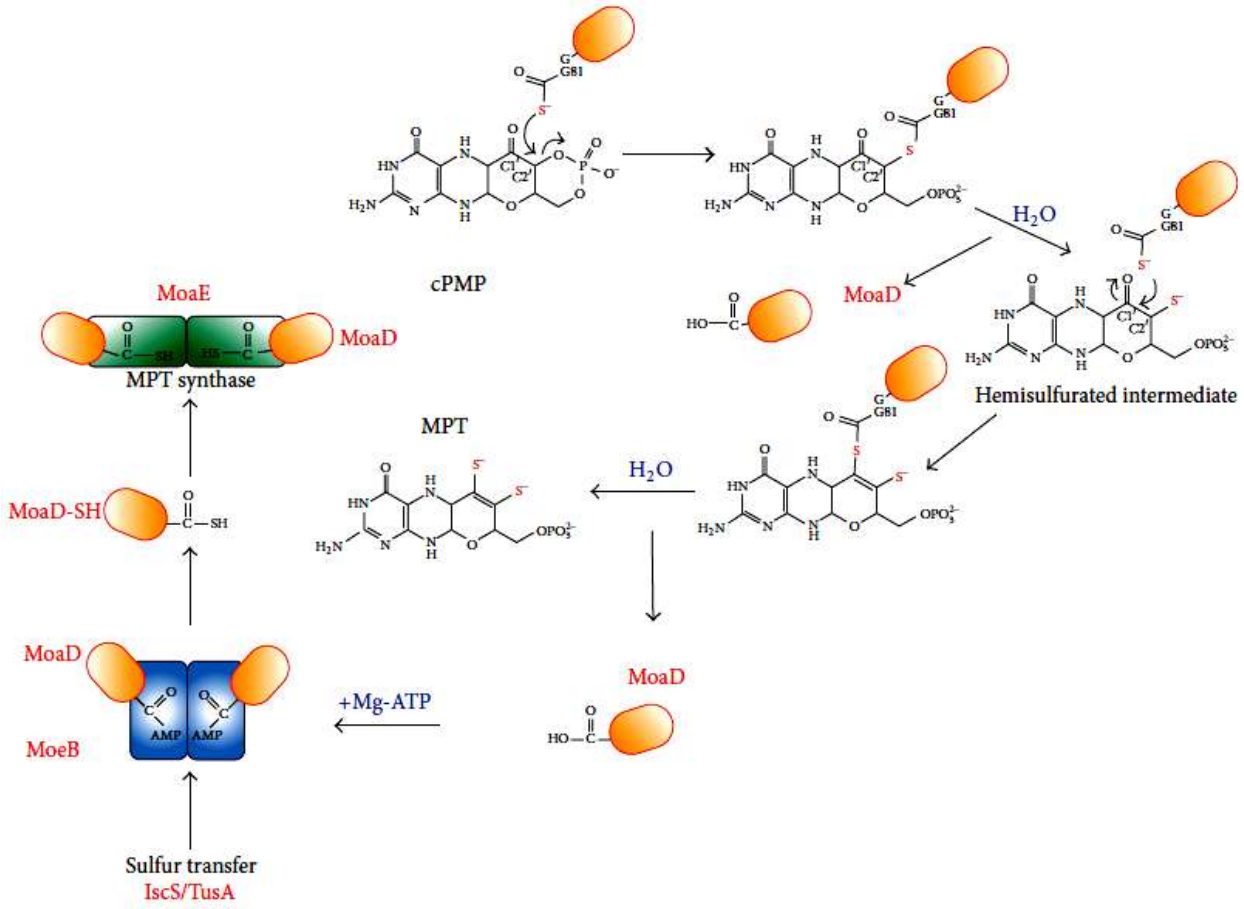
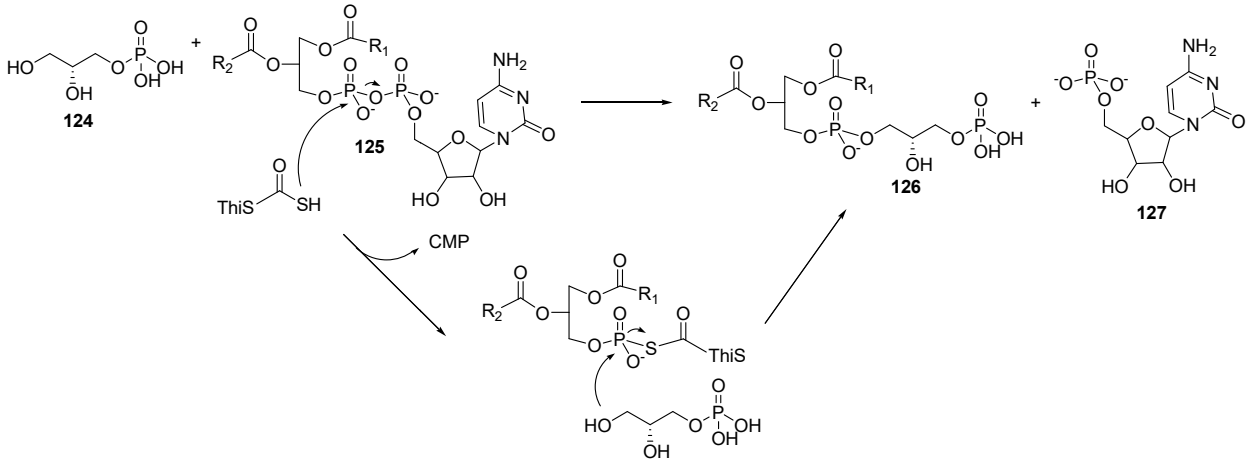


Figure 7.23: Proposed function of ThiS-COSH in CDP_OH_P and UvrABC system protein C catalyzed reaction. A) MoeE originally was found to catalyze the di-sulfur incorporation into the molybdopterin cofactor using MoaD as the sulfur donor. B) Proposed ThiS/MoaD-COSH involved mechanism in CDP_OH_P catalyzed reaction. C) Proposed mechanism of UvrABC system protein C cleavage DNA at both 3' and 5' end catalyzed by MoeE dimer using two molecules of ThiS/MoaD-COSHs.

B)



C)

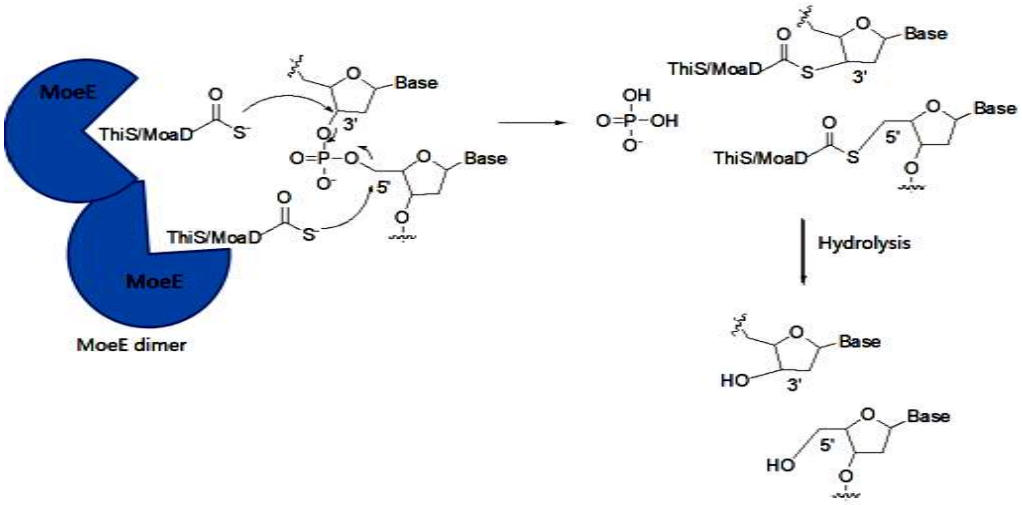


Figure 7.23: Continued

7.2.5 Thiamin phosphate synthase (ThiE)

ThiEs are in the thiamine gene clusters catalyzing the coupling of the thiamine thiazole part and pyrimidine part (Figure 7.26, the reaction in black). Here in over 100 cyanobacterial genomes (>30 genera), we found the conserved gene clusters with only ThiS and ThiE (examples in Figure 7.24). A predicted structure of ThiS in this cluster was also similar to the ubiquitin, though it only contained 1 α-helix but 5 β-sheets (usually it contains 2-3 α-helices and 5 β-sheets) (Figure 7.25). A proposed mechanism of the ThiE-ThiS mediated reaction was shown in Figure 7.26 in red. The new product **130** might be functioned as an antivitamin, or a ROS quencher. Some cyanobacterial genomes with this ThiS-ThiE gene cluster only contain one ThiE and one ThiS, and they are required for the thiamine biosynthesis. There may be a regulation machinery for the cyanobacteria to switch between the two functions if my proposed reaction does exist. The predicted ThiS-like protein structure was less similar to the ubiqitin structure compared with other ThiS-like proteins, maybe suggesting some unusual function of this type of ThiS-like proteins.

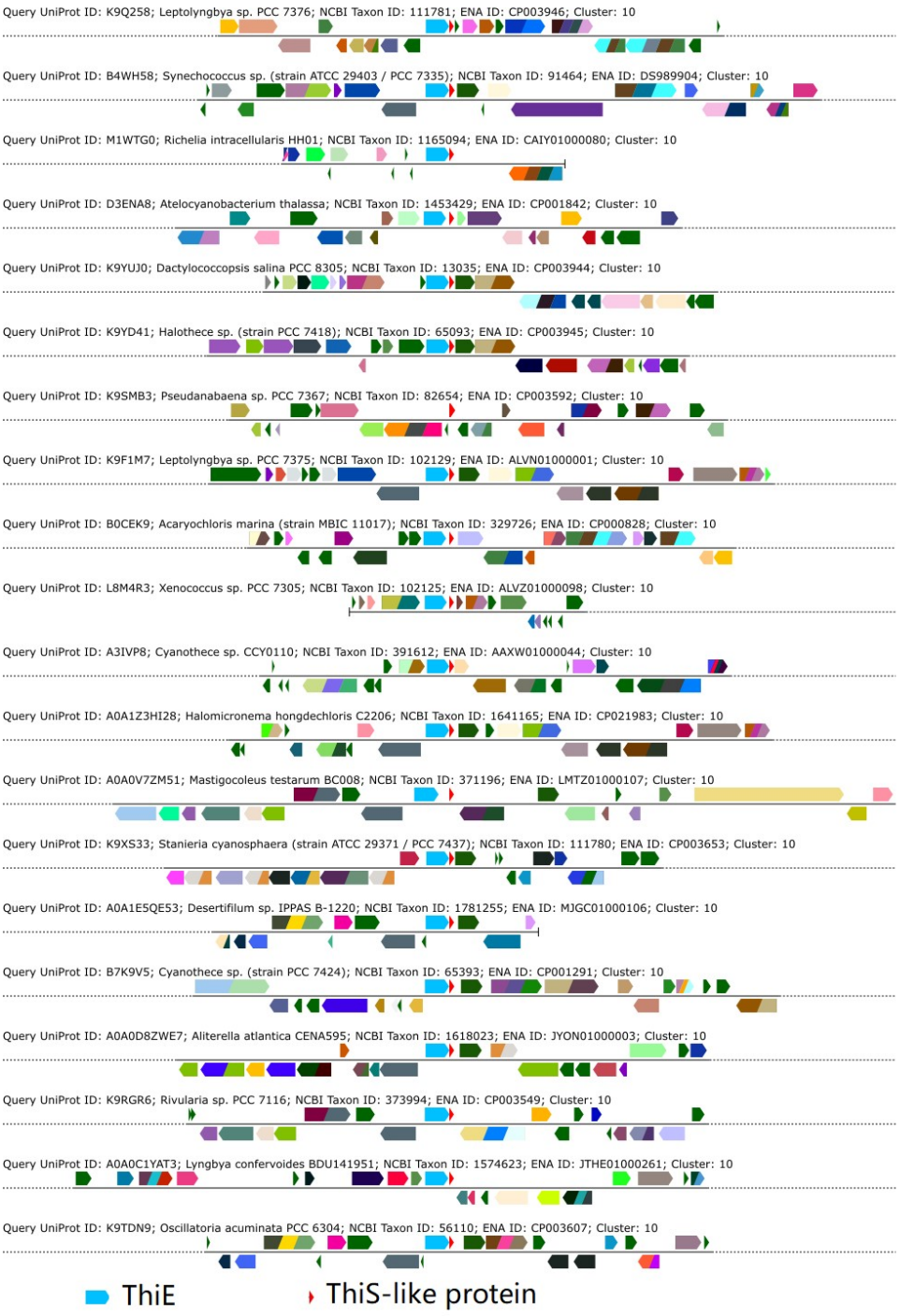


Figure 7.24: Examples of ThiE genes clusted with a conserved ThiS-like protein gene in various genera genomes. (Ligands of conserved genes were shown at the bottom) No other thiamine biosynthesis genes were found in the neighborhood of *thiS* gene (±10 genes).

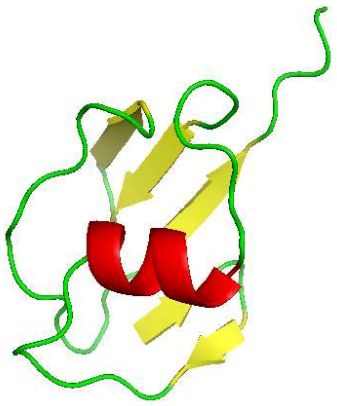


Figure 7.25: Predicted structure of a conserved ThiS-like protein next to ThiE (UniProtKB: K9F1M7). It showed the ubiquitin-like structure with a long C-terminal tail.

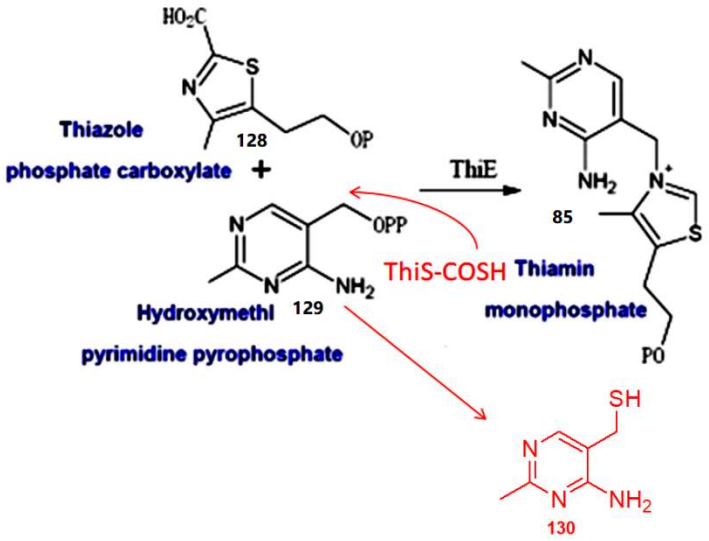


Figure 7.26: Proposed mechanism and product that ThiE catalyzed with the existence of ThiS-COSH. The reaction in black was the original ThiE catalyzed thiamin monophosphate formation, and the reaction in red was the hypothetical ThiE-ThiS coupled reaction.

7.2.6 ThiS-like protein involved in Chuangxinmycin biosynthesis—a pathway easy to be missed through pure bioinformatical searching

The gene cluster responsible for the biosynthesis of Chuangxinmycin has been reported (Figure 7.27).[73] This pathway has a ThiS-like protein. Bioinformatics search for this

gene clusters showed that only 5 genomes from 3 bacterial genera had the conserved gene cluster (Figure 7.28, red boxed). Other enzymes in this gene cluster are very commonly found in many different gene clusters. In this case, it is almost impossible to discover this new ThiS-involved biosynthesis pathway solely based on the bioinformatics analysis. The most reliable way is still the experimental detection of the related small metabolites, or the labeling of the trace amount of ThiS-COSH in the proteome range. Therefore, the most efficient way for new ThiS pathways searching is to combine both the computational biology and experimental biochemistry techniques to make more breakthroughs on discovering new sulfur chemistry in biosystems.

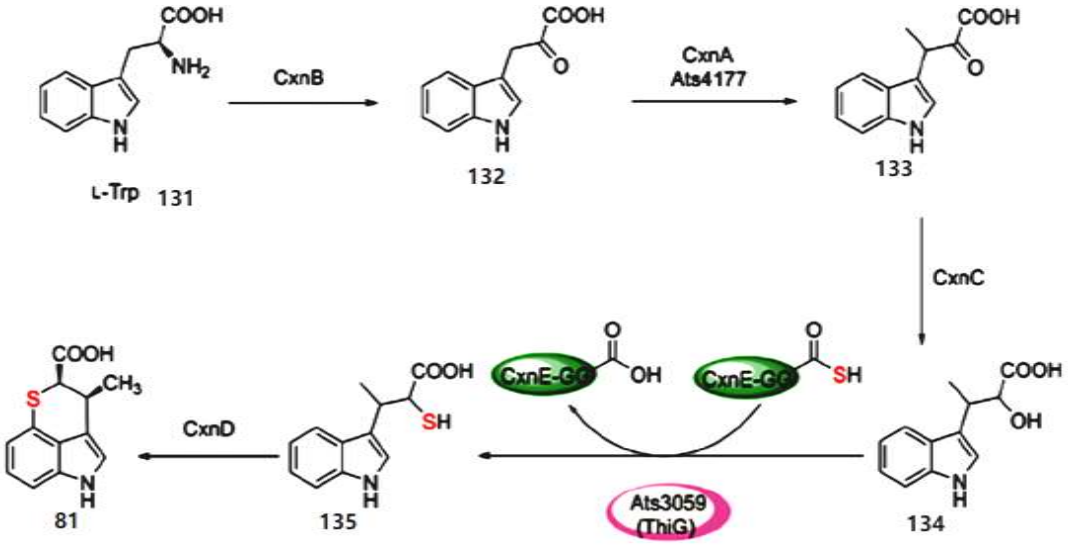


Figure 7.27: Proposed gene cluster involved in Chuangxinmycin biosynthesis.[73] CxnB: aminotransferase; CxnA: radical SAM methyltransferase; CxnC: reductase; ThiG: sulfur incorporation enzyme (ThiG was located elsewhere not in the gene cluster); CxnD: cytochrome P450. (See Figure 7.28 for a list of conserved Chuangxinmycin biosynthesis gene cluster among various genera)

Figure 7.28: Chuangxinmycin biosynthesis gene cluster was found conserved in only 5 genomes over 3 bacterial genera from an NCBI genome database searching.

7.3 Conclusion

We have surveyed the sequence similarity network (SSN) and the genome neighborhood network (GNN) of ThiS/MoaD sulfur carrier InterPro family (IPR003749) and found five highly conserved gene clusters (spanned over 30 genera of bacteria) that may lead to new sulfur chemistry in the biological systems. ThiS-COSH may be functioned as a general base catalyst, a sulfur donor, a thiocarbonate donor, an activation reagent (by forming thioesters) and an enzyme-catalyzed acylium ion generator (the second known case if it is true). Still, this pure "dry" searching may not be suitable to find new pathways only conserved in a narrow range of bacterial species and with all clustered enzymes commonly seen around the whole genome. A hybrid of powerful bioinformatics analysis and ultra sensitive/selective proteomic labeling/analysis tools and a diversity of robust biochemistry approaches is essential for current enzymology study.

7.4 **Materials and methods**

Bioinformatics study details: The ThiS-like protein sequence similarity networks (SSNs) and genome neighborhood networks (GNNs) were generated automatically from online EFI enzyme function initiative tools (https://efi.igb.illinois.edu/) using the Sulfur carrier ThiS/MoaD-like InterPro family with 38482 proteins matched (IPR003749). The generated networks were visualized using Cytoscape 3.4 software and analyzed following the EFI-EST and EFI-GNT online tutorial materials. The conserved gene cluster retrietment was also done using EFI-GNT tool (https://efi.igb.illinois.edu/efi-gnt/) and explored the gene cluster comparison plots via Inkscape software following the tutorials too. The predicted protein structures were generated from an online protein threading tool, Protein Homology/analogY Recognition Engine V 2.0

(Phrye 2) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). We thank Dr. Jamison P. Huddleston for providing me some informative trainings on using the SSN and GNN tools.

8. SUMMARY AND FUTURE PERSPECTIVE

Firstly, we discovered the first case of acylium ion intermediates formed in an enzyme (QDO) catalyzed reaction and were able to trap the acylium ion intermediates experimentally (Figure 8.1). Acylium ions are cations that usually can be efficiently trapped by solvent water. The native substrate was buried inside the QDO binding pocket which prevented the solvent water from reaching the acylium ion intermediates before they are converting to the final product (2). We used an unnatural substrate, 3,7-dihydroxyquinolin-4(1H)-one (**50**), to trap the first acylium ion. The 7-OH substitution of **50** may affect its binding and cause the first acylium ion intermediate partially extruding the binding pocket and being trapped by the solvent water. The W153G QDO mutant opened up the protein binding pocket further for the solvent water entering by which the second acylium ion was finally trapped. The solvent-trapping products of the acylium ion intermediates were observed when the reaction was performed in $^{18}O_2$ or $H_2{}^{18}O$ and analyzed by LC-MS. The ^{18}O labeling position of the solvent-trapping products was also determined by LC-MS/MS analysis. All the results were consistent with the proposed acylium ion mechanism. A feasible reaction energy profile was also calculated using DFT methods with a simplified system only including the substrate and the essential residues. MM methods were used to model the opened active site of W153G QDO mutant. In the future, experiments can be designed for acylium ion in situ trapping using stop-flow technique and kinetic study. Both the enzyme and the substrate may also be engineered further to serve this in situ trapping purpose. Since QDO catalyzes a robust and efficient CO releasing reaction, applications may be explored based on this feature, such as the development of biosensors and CO-related therapies. The photochemistry study of the substrate **1** may also provide some useful information on the

reaction mechanism study. With the successful trapping of the first case of enzyme-formed acylium ion intermediates, we are motivated in seeking for more enzymatic systems involved in the acylium ion formation. Computationally, we can further calculate the reaction energy profile in a big model including all binding pocket residues or even the whole protein using ONIOM (our own n-layered integrated molecular orbital and molecular mechanics).[106] This method is a hybrid of the DFT calculation at the active site (higher theory level layer) together and the MM optimization of the protein shell (lower theory level layer). These calculations will help us better understanding how the QDO enzyme stabilizes the intermediates and control the specificity of product formation.

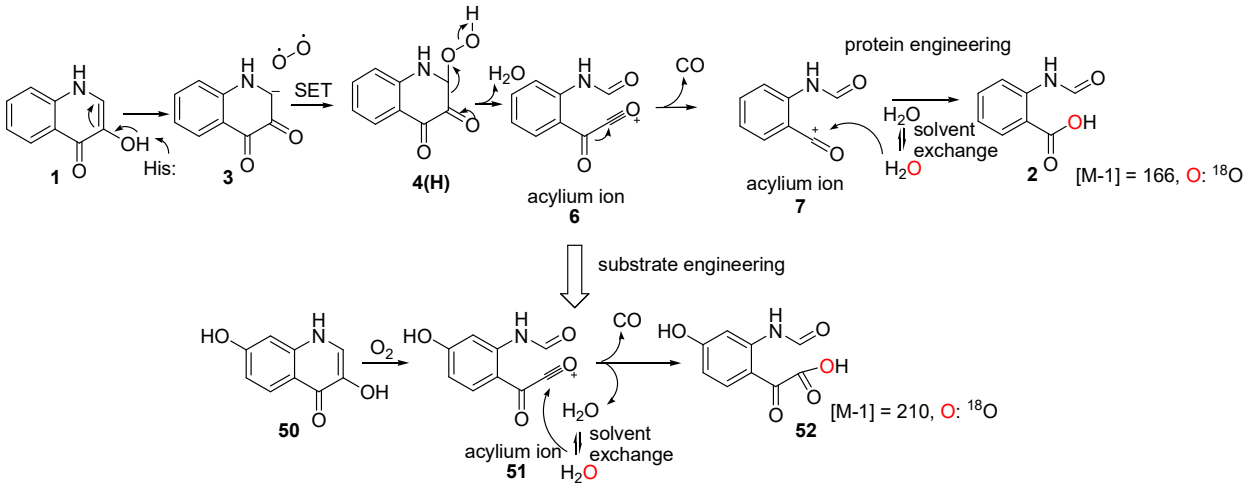


Figure 8.1: Proposed mechanism of 1-H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase (QDO) catalyzed reaction. The acylium ions **6** and **7** are experimentally trapped by $H_2{}^{18}O$ (^{18}O is in red).

Secondly, we figured out that the LRSA-lysine interaction might be a major cause for the sulfonyl azide reagent, LRSA (**90**), non-specific background labeling (Figure 8.2). The non-specific LRSA-lysine interaction was studied using the small molecule model reaction, the purified protein thiocarboxylate (ThiS-$COSH_{Tt}$), and the bacterial proteomes from *E. coli* and *Lc. lactis*. The small primary amine (2-methoxyethanamine, **97**) reacted much slower with LRSA

than the small thiocarboxylate (thioacetic acid, **93**) at neutral pH. In the proteome, the lysine amine might be affected by the surrounding amino acid residues and be more reactive for LRSA labeling. The lysine residues were much more abundant than the protein thiocarboxylates. Thus the non-specific LRSA-lysine interaction was interfering with protein thiocarboxylate labeling and sometimes caused "false positive" in the SDS-PAGE analysis. With the addition of NH_2OH (**99**) and heating for 20 min, the LRSA-protein thiocarboxylate adducts were decomposed and released free rhodamine residues while the LRSA-lysine adducts were stable. The LRSA (**90**) labeling time was optimized to 15 min for efficient and selective labeling the protein thiocarboxylates in *E. coli* and *Lc. Lactis* proteomes. To better eliminate the non-specific sulfonyl azide-lysine labeling for new protein thiocarboxylates screening from the environmental microbiome in the future, the lysine amine may be specifically labeled first with another reagent to block its nucleophilicity before LRSA (**90**) labeling. The lysine-block reagents should not react with thiocarboxylates but should react rapidly with amines in mild conditions. The possible candidates are succinimidyl esters, tetrafluorophenyl esters, and sulfodichlorophenol esters.

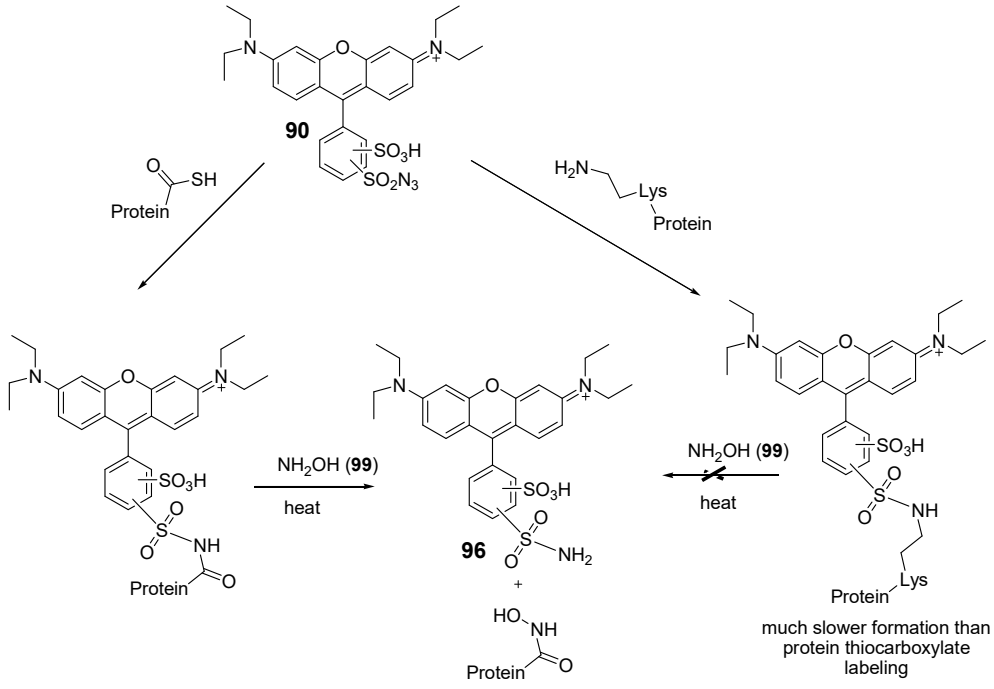


Figure 8.2: Scheme of LRSA (**90**) labeling protein thiocarboxylates and the non-specific lysine labeling.

We also applied the optimized the sulfonyl azide-protein thiocarboxylate labeling conditions to non-ThiS-like protein thiocarboxylates and small thiocarboxylates labeling. In LarA cofactor biosynthesis pathway, the aryl sulfonyl azide "click" reagents were used to selectively label the two protein thiocarboxylates (LarE-COSH, **14** and LarA-COSH, **18**) and two small molecule thiocarboxylate metabolites (**12** and **15**) separately from the bacteria proteomes and metabolomes (Figure 8.3). Protein thiocarboxylation used to be only found in ThiS-like sulfur carrier proteins. LarA-COSH and LarE-COSH were the first reported case of non-ThiS-like protein thiocarboxylates. The LarA cofactor biosynthesis gene clusters (*larABCE*) are widely distributed in bacterial species. Some stains contain the *lar* gene clusters but miss *larA* or more genes, which suggests proteins other than LarA may also use the larA cofactor for new functions. The sulfonyl azide based chemoproteomic labeling approaches can help us detect

new LarA-like protein thiocarboxylates in the bacterial proteomes for further biofunction characterizations. With the improved labeling selectivity and sensitivity, the sulfonyl azide reagents may have broad applications in discovering new protein thiocarboxylation sites other than ThiS-COSH or LarA-COSH from environmental microbiome.

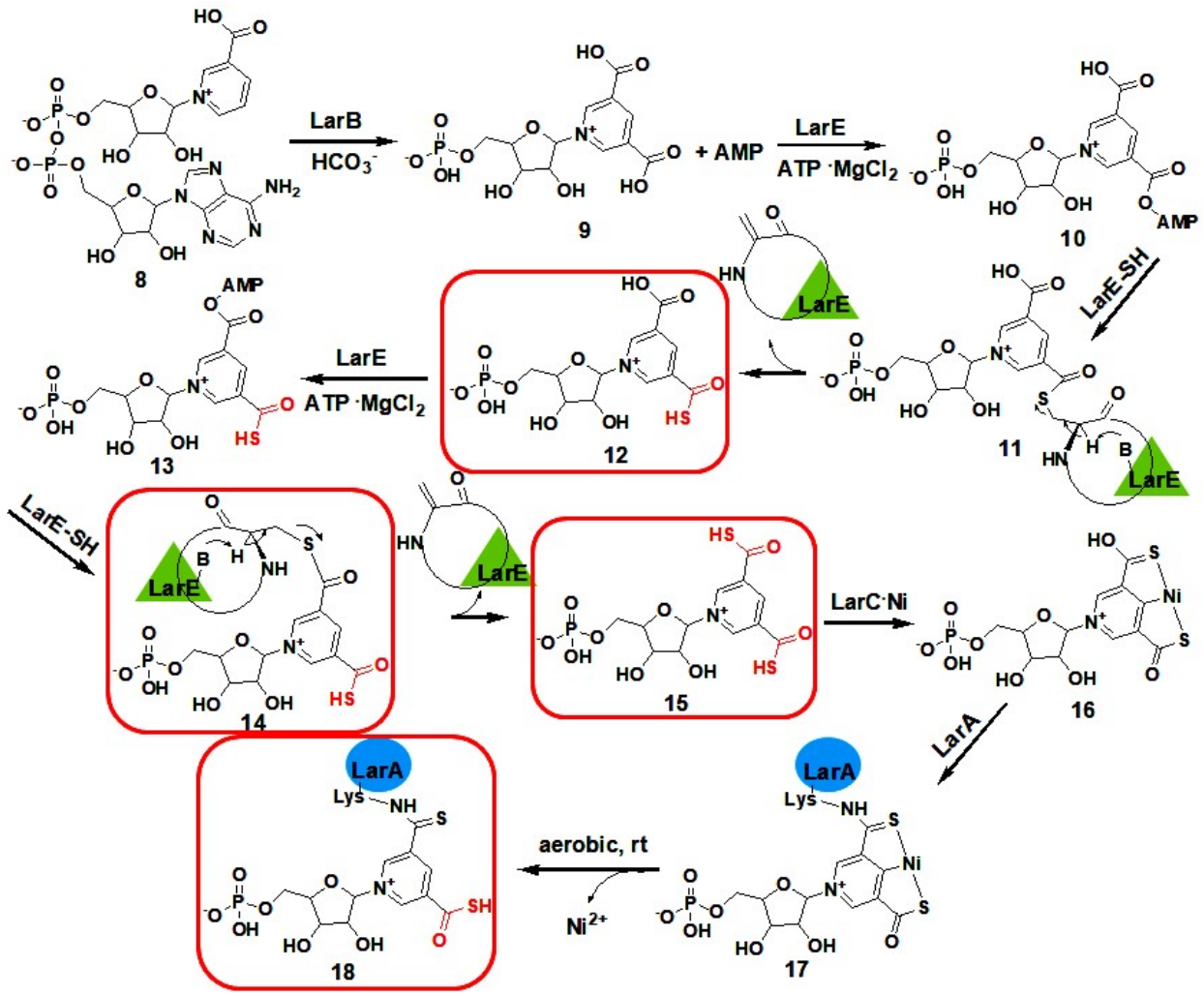


Figure 8.3: The LarA cofactor biosynthesis pathway. The protein thiocarboxylates (**14** and **18**) and small thiocarboxylate metabolites (**12** and **15**) were selectively labeled in the bacterial cell extract (in red boxes) using sulfonyl azide reagents.

Through a sequence similarity networks (SSNs) and genome neighborhood networks (GNNs) analysis using the EFI-EST web-tool (https://efi.igb.illinois.edu/efi-est/), I have found

that several protein families spanned many bacterial strains with a conservative gene neighbor of ThiS-like protein but no other thiamin, or molybdopterin biosynthesis gene around (within 20 gene neighbors). These ThiS-like proteins may participate in unknown natural product biosynthesis pathways from where some interesting new chemistry will be explored. A graduate student in our group is currently working on the mechanism study of a ThiS-like protein in Chuangxinmycin (**81**) biosynthesis pathway[73]. Both ThiS-like sulfur carrier proteins and LarA homologs are widespread in various bacteria strains[21,87]. The diversity of their gene context suggests the novel biosynthesis pathways that they might participate. Such bioinformatic analysis becomes our motivation in developing the robust protein thiocarboxylate chemical probes. Combination of the "wet" experimental measurements and "dry" computational bioinformatic data mining will increase our efficiency in discovering new sulfur biochemistry in the future.

REFERENCES

(1) Lewis, C. A.; Wolfenden, R. Uroporphyrinogen Decarboxylation as a Benchmark for the Catalytic Proficiency of Enzymes. *Proc. Natl. Acad. Sci.* **2008**, *105* (45), 17328.

(2) Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. Mechanisms and Free Energies of Enzymatic Reactions. *Chem. Rev.* **2006**, *106* (8), 3188–3209.

(3) Tantillo, D. J. The Carbocation Continuum in Terpene Biosynthesis—Where Are the Secondary Cations? *Chem. Soc. Rev.* **2010**, *39* (8), 2847–2854.

(4) Tang, M.-C.; Zou, Y.; Watanabe, K.; Walsh, C. T.; Tang, Y. Oxidative Cyclization in Natural Product Biosynthesis. *Chem. Rev.* **2017**, *117* (8), 5226–5333.

(5) Lietzan, A. D.; St. Maurice, M. Functionally Diverse Biotin-Dependent Enzymes with Oxaloacetate Decarboxylase Activity. *Arch. Biochem. Biophys.* **2014**, *544*, 75–86.

(6) Pohl, M.; Sprenger, G. A.; Müller, M. A New Perspective on Thiamine Catalysis. *Curr. Opin. Biotechnol.* **2004**, *15* (4), 335–342.

(7) Mak, P. J.; Thammawichai, W.; Wiedenhoeft, D.; Kincaid, J. R. Resonance Raman Spectroscopy Reveals PH-Dependent Active Site Structural Changes of Lactoperoxidase Compound 0 and Its Ferryl Heme O–O Bond Cleavage Products. *J. Am. Chem. Soc.* **2015**, *137* (1), 349–361.

(8) Mitić, N.; Clay, M. D.; Saleh, L.; Bollinger, J. Martin; Solomon, E. I. Spectroscopic and Electronic Structure Studies of Intermediate X in Ribonucleotide Reductase R2 and Two Variants: A Description of the FeIV-Oxo Bond in the FeIII−O−FeIV Dimer. *J. Am. Chem. Soc.* **2007**, *129* (29), 9049–9065.

(9) Mak, P. J.; Denisov, I. G. Spectroscopic Studies of the Cytochrome P450 Reaction Mechanisms. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2018**, *1866* (1), 178–204.

(10) Lipscomb, J. D. Mechanism of Extradiol Aromatic Ring-Cleaving Dioxygenases. *Curr. Opin. Struct. Biol.* **2008**, *18* (6), 644–649.

(11) Sicoli, G.; Mouesca, J.-M.; Zeppieri, L.; Amara, P.; Martin, L.; Barra, A.-L.; Fontecilla-Camps, J. C.; Gambarelli, S.; Nicolet, Y. Fine-Tuning of a Radical-Based Reaction by Radical S-Adenosyl-L-Methionine Tryptophan Lyase. *Science* **2016**, *351* (6279), 1320–1323.

(12) Bhandari, D. M.; Fedoseyenko, D.; Begley, T. P. Mechanistic Studies on Tryptophan Lyase (NosL): Identification of Cyanide as a Reaction Product. *J. Am. Chem. Soc.* **2018**, *140* (2), 542–545.

(13) Huang, H.; Chang, W.-C.; Lin, G.-M.; Romo, A.; Pai, P.-J.; Russell, W. K.; Russell, D. H.; Liu, H.-W. Mechanistic Consequences of Chiral Radical Clock Probes: Analysis of the Mononuclear Non-Heme Iron Enzyme HppE with 2-Hydroxy-3-Methylenecyclopropyl Radical Clock Substrates. *J. Am. Chem. Soc.* **2014**, *136* (8), 2944–2947.

(14) Epand, R. M.; Wilson, I. B. Evidence for the Formation of Hippuryl Chymotrypsin during the Hydrolysis of Hippuric Acid Esters. *J. Biol. Chem.* **1963**, *238* (5), 1718–1723.

(15) Bhandari, D. M.; Fedoseyenko, D.; Begley, T. P. Tryptophan Lyase (NosL): A Cornucopia of 5′-Deoxyadenosyl Radical Mediated Transformations. *J. Am. Chem. Soc.* **2016**, *138* (50), 16184–16187.

(16) Maddur, A. A.; Swanson, R.; Izaguirre, G.; Gettins, P. G. W.; Olson, S. T. Kinetic Intermediates En Route to the Final Serpin-Protease Complex STUDIES OF COMPLEXES OF A1-PROTEASE INHIBITOR WITH TRYPSIN. *J. Biol. Chem.* **2013**, *288* (44), 32020–32035.

(17) Bushnell, E. A. C.; Berryman, V. E. J.; Gauld, J. W.; Boyd, R. J. Chapter Six - The Importance of the MM Environment and the Selection of the QM Method in QM/MM Calculations: Applications to Enzymatic Reactions. In *Advances in Protein Chemistry and Structural Biology*; Karabencheva-Christova, T., Ed.; Combined Quantum Mechanical and Molecular Mechanical Modelling of Biomolecular Interactions; Academic Press, 2015; Vol. 100, pp 153–185.

(18) Romero, J. M.; Martin, M.; Ramirez, C. L.; Dumas, V. G.; Marti, M. A. Chapter Two - Efficient Calculation of Enzyme Reaction Free Energy Profiles Using a Hybrid Differential Relaxation Algorithm: Application to Mycobacterial Zinc Hydrolases. In *Advances in Protein Chemistry and Structural Biology*; Karabencheva-Christova, T., Ed.; Combined Quantum Mechanical and Molecular Mechanical Modelling of Biomolecular Interactions; Academic Press, 2015; Vol. 100, pp 33–65.

(19) Structural basis for cofactor-independent dioxygenation of N-heteroaromatic compounds at the α/β-hydrolase fold | PNAS http://www.pnas.org.ezproxy.library.tamu.edu/content/107/2/657.long (accessed Apr 24, 2018).

(20) Mehta, A. P.; Abdelwahed, S. H.; Fenwick, M. K.; Hazra, A. B.; Taga, M. E.; Zhang, Y.; Ealick, S. E.; Begley, T. P. Anaerobic 5-Hydroxybenzimidazole Formation from Aminoimidazole Ribotide: An Unanticipated Intersection of Thiamin and Vitamin B12 Biosynthesis https://pubs.acs.org/doi/abs/10.1021/jacs.5b03576 (accessed Apr 23, 2018).

(21) Krishnamoorthy, K.; Begley, T. P. Reagent for the Detection of Protein Thiocarboxylates in the Bacterial Proteome: Lissamine Rhodamine B Sulfonyl Azide. *J. Am. Chem. Soc.* **2010**, *132* (33), 11608–11612.

(22) Jessani, N.; Cravatt, B. F. The Development and Application of Methods for Activity-Based Protein Profiling. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 54–59.

(23) Saghatelian, A.; Trauger, S. A.; Want, E. J.; Hawkins, E. G.; Siuzdak, G.; Cravatt, B. F. Assignment of Endogenous Substrates to Enzymes by Global Metabolite Profiling †. *Biochemistry (Mosc.)* **2004**, *43* (45), 14332–14339.

(24) Petras, D.; Jarmusch, A. K.; Dorrestein, P. C. From Single Cells to Our Planet—Recent Advances in Using Mass Spectrometry for Spatially Resolved Metabolomics. *Curr. Opin. Chem. Biol.* **2017**, *36*, 24–31.

(25) Huan, T.; Forsberg, E. M.; Rinehart, D.; Johnson, C. H.; Ivanisevic, J.; Benton, H. P.; Fang, M.; Aisporna, A.; Hilmers, B.; Poole, F. L.; et al. Systems Biology Guided by XCMS Online Metabolomics. *Nat. Methods* **2017**, *14* (5), 461–462.

(26) Protein Thiocarboxylate-Dependent Methionine Biosynthesis in Wolinella succinogenes - Journal of the American Chemical Society (ACS Publications) https://pubs.acs.org.ezproxy.library.tamu.edu/doi/abs/10.1021/ja107424t (accessed Apr 23, 2018).

(27) Desguin, B.; Zhang, T.; Soumillion, P.; Hols, P.; Hu, J.; Hausinger, R. P. METALLOPROTEINS. A Tethered Niacin-Derived Pincer Complex with a Nickel-Carbon Bond in Lactate Racemase. *Science* **2015**, *349* (6243), 66–69.

(28) Groves, J. K. The Friedel–Crafts Acylation of Alkenes. *Chem. Soc. Rev.* **1972**, *1* (1), 73–97.

(29) Shi, H.; Wang, Y.; Hua, R. Acid-Catalyzed Carboxylic Acid Esterification and Ester Hydrolysis Mechanism: Acylium Ion as a Sharing Active Intermediate via a Spontaneous Trimolecular Reaction Based on Density Functional Theory Calculation and Supported by Electrospray Ionization-Mass Spectrometry. *Phys. Chem. Chem. Phys.* **2015**, *17* (45), 30279–30291.

(30) Lindner, E. The Behavior of Acid Halides Toward Lewis Acids and Lewis Bases. *Angew. Chem. Int. Ed. Engl.* **1970**, *9* (2), 114–123.

(31) Wedler, H.; Pemberton, R.; Tantillo, D. Carbocations and the Complex Flavor and Bouquet of Wine: Mechanistic Aspects of Terpene Biosynthesis in Wine Grapes. *Molecules* **2015**, *20* (6), 10781–10792.

(32) Olah, G. Acylium Ions (Acyl Cations). In *Carbonium ions*; 1976; pp 2049–2133.

(33) Moraes, L. A. B.; Eberlin, M. N. Acyclic Distonic Acylium Ions: Dual Free Radical and Acylium Ion Reactivity in a Single Molecule. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (8), 697–704.

(34) Lyall, C. L.; Sato, M.; Uosis-Martin, M.; Asghar, S. F.; Jones, M. D.; Williams, I. H.; Lewis, S. E. C–H Functionalization of Sp3 Centers with Aluminum: A Computational and Mechanistic Study of the Baddeley Reaction of Decalin. *J. Am. Chem. Soc.* **2014**, *136* (39), 13745–13753.

(35) *Active Oxygen in Biochemistry*; Valentine, J., Foote, C. S., Greenberg, A., Liebman, J. F., Eds.; Structure Energetics and Reactivity in Chemistry Series; Springer Netherlands, 1995.

(36) Fetzner, S.; Steiner, R. A. Cofactor-Independent Oxidases and Oxygenases. *Appl. Microbiol. Biotechnol.* **2010**, *86* (3), 791–804.

(37) BOTT, G.; SCHMIDT, M.; ROMMEL, T. O.; LINGENS, F. Microbial Metabolism of Quinoline and Related Compounds. V. Degradation of 1H-4-Oxoquinoline by Pseudomonas Putida 33/1. *Biol. Chem. Hoppe. Seyler* **2009**, *371* (2), 999–1004.

(38) Max, N.; Betz, A.; Facey, S.; Lingens, F.; Hauer, B.; Fetzner, S. Cloning, Sequence Analysis, and Expression of the Pseudomonas Putida 33/1 1H-3-Hydroxy-4-Oxoquinoline 2,4-Dioxygenase Gene, Encoding a Carbon Monoxide Forming Dioxygenase1The Sequence of the 1H-3-Hydroxy-4-Oxoquinoline 2,4-Dioxygenase (Qdo) Gene Qdo Reported in This Paper Has Been Deposited in the EMBL Data Library under the Accession Number Y14779.1. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **1999**, *1431* (2), 547–552.

(39) Mann, B. E. Carbon Monoxide: An Essential Signalling Molecule. In *Medicinal Organometallic Chemistry*; Topics in Organometallic Chemistry; Springer, Berlin, Heidelberg, 2010; pp 247–285.

(40) Ng, P. C. Y.; Long, B.; Koyfman, A. Clinical Chameleons: An Emergency Medicine Focused Review of Carbon Monoxide Poisoning. *Intern. Emerg. Med.* **2018**, *13* (2), 223–229.

(41) Peers, C.; Steele, D. S. Carbon Monoxide: A Vital Signalling Molecule and Potent Toxin in the Myocardium. *J. Mol. Cell. Cardiol.* **2012**, *52* (2), 359–365.

(42) Wang, M.; Liao, W. Carbon Monoxide as a Signaling Molecule in Plants. *Front. Plant Sci.* **2016**, *7*.

(43) Allpress, C. J.; Berreau, L. M. Oxidative Aliphatic Carbon–Carbon Bond Cleavage Reactions. *Coord. Chem. Rev.* **2013**, *257* (21), 3005–3029.

(44) Sun, Y.-J.; Huang, Q.-Q.; Zhang, J.-J. A Series of NiII-Flavonolate Complexes as Structural and Functional ES (Enzyme-Substrate) Models of the NiII-Containing Quercetin 2,3-Dioxygenase. *Dalton Trans.* **2014**, *43* (17), 6480–6489.

(45) Balogh-Hergovich, É.; Kaizer, J.; Speier, G.; Fülöp, V.; Párkányi, L. Quercetin 2,3-Dioxygenase Mimicking Ring Cleavage of the Flavonolate Ligand Assisted by Copper. Synthesis and Characterization of Copper(I) Complexes [Cu(PPh3)2(Fla)] (Fla = Flavonolate) and [Cu(PPh3)2(O-Bs)] (O-Bs = O-Benzoylsalicylate). *Inorg. Chem.* **1999**, *38* (17), 3787–3795.

(46) Sun, Y.-J.; Huang, Q.-Q.; Zhang, J.-J. Series of Structural and Functional Models for the ES (Enzyme–Substrate) Complex of the Co(II)-Containing Quercetin 2,3-Dioxygenase. *Inorg. Chem.* **2014**, *53* (6), 2932–2942.

(47) Bauer, I.; Max, N.; Fetzner, S.; Lingens, F. 2,4-Dioxygenases Catalyzing N-Heterocyclic-Ring Cleavage and Formation of Carbon Monoxide. *Eur. J. Biochem.* **1996**, *240* (3), 576–583.

(48) Chien, T. C.; Dias, L. G.; Arantes, G. M.; Santos, L. G. C.; Triboni, E. R.; Bastos, E. L.; Politi, M. J. 1-(2-Quinolyl)-2-Naphthol: A New Intra-Intermolecular Photoacid–Photobase Molecule. *J. Photochem. Photobiol. Chem.* **2008**, *194* (1), 37–48.

(49) Evans, D. J.; Eastwood, F. W. Synthesis of an Arylhydroxytetronimide and of 3-Hydroxy-4(1H)-Quinolone Derivatives. *Aust. J. Chem.* **1974**, *27* (3), 537–542.

(50) Szczepankiewicz, W.; Suwinski, J. One-Pot Synthesis of 3-(2-Cyanophenyl)Quinazolin-4(3H)-One. *Chem. Heterocycl. Compd.* **2000**, *36* (7), 809–810.

(51) Fun, H.-K.; Goh, J. H.; Chandrakantha, B.; Isloor, A. M. 1-Dibromomethyl-4-Methoxy-2-Nitrobenzene. *Acta Crystallogr. Sect. E Struct. Rep. Online* **2009**, *65* (9), o2193–o2194.

(52) Gaillard, P.; Pomel, V.; Jeanclaude-Etter, I.; Dorbais, J.; Klicic, J.; Montagne, C. Quinoxaline Compounds and Use Thereof. WO2008101979 (A1), August 28, 2008.

(53) Geierstanger, B.; Ou, W.; Cellitti, S.; Uno, T.; Crossgrove, T.; Chiu, H.-P.; Grunewald, J.; Hao, X. Biosynthetically Generated Pyrroline-Carboxy-Lysine and Site Specific Protein

Modifications Via Chemical Derivatization of Pyrroline-Carboxy-Lysine and Pyrrolysine Residues. WO2010048582 (A1), April 29, 2010.

(54) Lewars, E. G. An Outline of What Computational Chemistry Is All About. In *Computational Chemistry*; Springer, Dordrecht, 2011; pp 1–7.

(55) Jones, R. O. Density Functional Theory: Its Origins, Rise to Prominence, and Future. *Rev. Mod. Phys.* **2015**, *87* (3), 897–923.

(56) Stewart, J. J. P. Semiempirical Molecular Orbital Methods. In *Reviews in Computational Chemistry*; Wiley-Blackwell, 2007; pp 45–81.

(57) Lewars, E. G. Density Functional Calculations. In *Computational Chemistry*; Springer, Dordrecht, 2011; pp 445–519.

(58) Lewars, E. G. Molecular Mechanics. In *Computational Chemistry*; Springer, Dordrecht, 2011; pp 45–83.

(59) Rubenstein, L. A.; Lanzara, R. G. Activation of G Protein-Coupled Receptors Entails Cysteine Modulation of Agonist Binding. *J. Mol. Struct. THEOCHEM* **1998**, *430*, 57–71.

(60) Rubenstein, L. A.; Zauhar, R. J.; Lanzara, R. G. Molecular Dynamics of a Biophysical Model for B2-Adrenergic and G Protein-Coupled Receptor Activation. *J. Mol. Graph. Model.* **2006**, *25* (4), 396–409.

(61) Hernández-Ortega, A.; Quesne, M. G.; Bui, S.; Heyes, D. J.; Steiner, R. A.; Scrutton, N. S.; de Visser, S. P. Catalytic Mechanism of Cofactor-Free Dioxygenases and How They Circumvent Spin-Forbidden Oxygenation of Their Substrates. *J. Am. Chem. Soc.* **2015**, *137* (23), 7474–7487.

(62) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.

(63) Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F - Clark - 1983 - Journal of Computational Chemistry

- Wiley Online Library https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540040303 (accessed Apr 25, 2018).

(64) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent Molecular Orbital Methods 25. Supplementary Functions for Gaussian Basis Sets. *J. Chem. Phys.* **1984**, *80* (7), 3265–3269.

(65) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72* (1), 650–654.

(66) Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions - The Journal of Physical Chemistry B (ACS Publications) https://pubs.acs.org/doi/abs/10.1021/jp810292n (accessed Apr 25, 2018).

(67) Quantum Mechanical Continuum Solvation Models - Chemical Reviews (ACS Publications) https://pubs.acs.org/doi/abs/10.1021/cr9904009 (accessed Apr 25, 2018).

(68) Gaussian 09 Citation | Gaussian.com http://gaussian.com/g09citation/ (accessed Apr 25, 2018).

(69) Leidel, S.; Pedrioli, P. G. A.; Bucher, T.; Brost, R.; Costanzo, M.; Schmidt, A.; Aebersold, R.; Boone, C.; Hofmann, K.; Peter, M. Ubiquitin-Related Modifier Urm1 Acts as a Sulphur Carrier in Thiolation of Eukaryotic Transfer RNA. *Nature* **2009**, *458* (7235), 228–232.

(70) Begley, T. P.; Xi, J.; Kinsland, C.; Taylor, S.; McLafferty, F. The Enzymology of Sulfur Activation during Thiamin and Biotin Biosynthesis. 7.

(71) Mueller, E. G. Trafficking in Persulfides: Delivering Sulfur in Biosynthetic Pathways. *Nat. Chem. Biol.* **2006**, *2* (4), 185–194.

(72) Burns, K. E.; Baumgart, S.; Dorrestein, P. C.; Zhai, H.; McLafferty, F. W.; Begley, T. P. Reconstitution of a New Cysteine Biosynthetic Pathway in Mycobacterium Tuberculosis. *J. Am. Chem. Soc.* **2005**, *127* (33), 11602–11603.

(73) Shi, Y.; Jiang, Z.; Li, X.; Zuo, L.; Lei, X.; Yu, L.; Wu, L.; Jiang, J.; Hong, B. Biosynthesis of Antibiotic Chuangxinmycin from Actinoplanes Tsinanensis. *Acta Pharm. Sin. B* **2018**, *8* (2), 283–294.

(74) Leimkühler, S.; Wuebbens, M. M.; Rajagopalan, K. V. Characterization of Escherichia Coli MoeB and Its Involvement in the Activation of Molybdopterin Synthase for the Biosynthesis of the Molybdenum Cofactor. *J. Biol. Chem.* **2001**, *276* (37), 34695–34701.

(75) Dorrestein, P. C.; Zhai, H.; McLafferty, F. W.; Begley, T. P. The Biosynthesis of the Thiazole Phosphate Moiety of Thiamin: The Sulfur Transfer Mediated by the Sulfur Carrier Protein ThiS. *Chem. Biol.* **2004**, *11* (10), 1373–1381.

(76) Chen, M.; Narai, S.; Omura, N.; Shigi, N.; Chimnaronk, S.; Tanaka, Y.; Yao, M. Crystallographic Study of the 2-Thioribothymidine-Synthetic Complex TtuA–TtuB from Thermus Thermophilus. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2016**, *72* (10), 777–781.

(77) Godert, A. M.; Jin, M.; McLafferty, F. W.; Begley, T. P. Biosynthesis of the Thioquinolobactin Siderophore: An Interesting Variation on Sulfur Transfer. *J. Bacteriol.* **2007**, *189* (7), 2941–2944.

(78) Sasaki, E.; Zhang, X.; Sun, H. G.; Lu, M.-Y. J.; Liu, T.; Ou, A.; Li, J.; Chen, Y.; Ealick, S. E.; Liu, H. Exploiting Sulphur-Carrier Proteins from Primary Metabolism for 2-Thiosugar Biosynthesis. *Nature* **2014**, *510* (7505), 427–431.

(79) Kolakowski, R. V.; Shangguan, N.; Sauers, R. R.; Williams, L. J. Mechanism of Thio Acid/Azide Amidation. *J. Am. Chem. Soc.* **2006**, *128* (17), 5695–5702.

(80) Merkx, R.; Brouwer, A. J.; Rijkers, D. T. S.; Liskamp, R. M. J. Highly Efficient Coupling of β-Substituted Aminoethane Sulfonyl Azides with Thio Acids, toward a New Chemical Ligation Reaction. *Org. Lett.* **2005**, *7* (6), 1125–1128.

(81) Shangguan, N.; Katukojvala, S.; Greenberg, R.; Williams, L. J. The Reaction of Thio Acids with Azides: A New Mechanism and New Synthetic Applications. *J. Am. Chem. Soc.* **2003**, *125* (26), 7754–7755.

(82) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40* (11), 2004–2021.

(83) Kolb, H. C.; Sharpless, K. B. The Growing Impact of Click Chemistry on Drug Discovery. *Drug Discov. Today* **2003**, *8* (24), 1128–1137.

(84) Agard, N. J.; Baskin, J. M.; Prescher, J. A.; Lo, A.; Bertozzi, C. R. A Comparative Study of Bioorthogonal Reactions with Azides. *ACS Chem. Biol.* **2006**, *1* (10), 644–648.

(85) Zhang, X.; Li, F.; Lu, X.-W.; Liu, C.-F. Protein C-Terminal Modification through Thioacid/Azide Amidation. *Bioconjug. Chem.* **2009**, *20* (2), 197–200.

(86) Kulikova, V. V.; Zakomirdina, L. N.; Bazhulina, N. P.; Dementieva, I. S.; Faleev, N. G.; Gollnick, P. D.; Demidkina, T. V. Role of Arginine 226 in the Mechanism of Tryptophan Indole-Lyase from Proteus Vulgaris. *Biochem. Mosc.* **2003**, *68* (11), 1181–1188.

(87) Desguin, B.; Goffin, P.; Viaene, E.; Kleerebezem, M.; Martin-Diaconescu, V.; Maroney, M. J.; Declercq, J.-P.; Soumillion, P.; Hols, P. Lactate Racemase Is a Nickel-Dependent Enzyme Activated by a Widespread Maturation System. *Nat. Commun.* **2014**, *5*, 3615.

(88) Desguin, B.; Soumillion, P.; Hols, P.; Hausinger, R. P. Nickel-Pincer Cofactor Biosynthesis Involves LarB-Catalyzed Pyridinium Carboxylation and LarE-Dependent Sacrificial Sulfur Insertion. *Proc. Natl. Acad. Sci.* **2016**, *113* (20), 5598–5603.

(89) Fellner, M.; Desguin, B.; Hausinger, R. P.; Hu, J. Structural Insights into the Catalytic Mechanism of a Sacrificial Sulfur Insertase of the N-Type ATP Pyrophosphatase Family, LarE. *Proc. Natl. Acad. Sci.* **2017**, *114* (34), 9074–9079.

(90) van der Greef Jan; Smilde Age K. Symbiosis of Chemometrics and Metabolomics: Past, Present, and Future. *J. Chemom.* **2006**, *19* (5-7), 376–386.

(91) Sasaki, E.; Liu, H. Mechanistic Studies of the Biosynthesis of 2-Thiosugar: Evidence for the Formation of an Enzyme-Bound 2-Ketohexose Intermediate in BexX-Catalyzed Reaction. *J. Am. Chem. Soc.* **2010**, *132* (44), 15544–15546.

(92) Mukund, S.; Adams, M. W. The Novel Tungsten-Iron-Sulfur Protein of the Hyperthermophilic Archaebacterium, Pyrococcus Furiosus, Is an Aldehyde Ferredoxin

Oxidoreductase. Evidence for Its Participation in a Unique Glycolytic Pathway. *J. Biol. Chem.* **1991**, *266* (22), 14208–14216.

(93) Chan, M. K.; Mukund, S.; Kletzin, A.; Adams, M. W.; Rees, D. C. Structure of a Hyperthermophilic Tungstopterin Enzyme, Aldehyde Ferredoxin Oxidoreductase. *Science* **1995**, *267* (5203), 1463–1469.

(94) Hu, Y.; Faham, S.; Roy, R.; Adams, M. W.; Rees, D. C. Formaldehyde Ferredoxin Oxidoreductase from Pyrococcus Furiosus: The 1.85 A Resolution Crystal Structure and Its Mechanistic Implications. *J. Mol. Biol.* **1999**, *286* (3), 899–914.

(95) Bevers, L. E.; Bol, E.; Hagedoorn, P.-L.; Hagen, W. R. WOR5, a Novel Tungsten-Containing Aldehyde Oxidoreductase from Pyrococcus Furiosus with a Broad Substrate Specificity. *J. Bacteriol.* **2005**, *187* (20), 7056–7061.

(96) Romão, M. J. Molybdenum and Tungsten Enzymes: A Crystallographic and Mechanistic Overview. *Dalton Trans. Camb. Engl. 2003* **2009**, No. 21, 4053–4068.

(97) Leimkühler, S. The Biosynthesis of the Molybdenum Cofactor in Escherichia coli and Its Connection to FeS Cluster Assembly and the Thiolation of tRNA https://www.hindawi.com/journals/ab/2014/808569/ (accessed Jul 3, 2018).

(98) Sugimoto, H.; Tano, H.; Tajima, R.; Miyake, H.; Tsukube, H.; Ohi, H.; Itoh, S. In Situ Generation of Oxo−sulfidobis(Dithiolene)Tungsten(VI) Complexes: Active-Site Models for the Aldehyde Ferredoxin Oxidoreductase Family of Tungsten Enzymes. *Inorg. Chem.* **2007**, *46* (21), 8460–8462.

(99) Hu, Y.; Faham, S.; Roy, R.; Adams, M. W.; Rees, D. C. Formaldehyde Ferredoxin Oxidoreductase from Pyrococcus Furiosus: The 1.85 A Resolution Crystal Structure and Its Mechanistic Implications. *J. Mol. Biol.* **1999**, *286* (3), 899–914.

(100) Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical S - Adenosylmethionine Enzymes. *Chem. Rev.* **2014**, *114* (8), 4229–4317.

(101) Fabrega, C.; Hausmann, S.; Shen, V.; Shuman, S.; Lima, C. D. Structure and Mechanism of MRNA Cap (Guanine-N7) Methyltransferase. *Mol. Cell* **2004**, *13* (1), 77–89.

(102) Attieh, J. M.; Hanson, A. D.; Saini, H. S. Purification and Characterization of a Novel Methyltransferase Responsible for Biosynthesis of Halomethanes and Methanethiol in Brassica Oleracea. *J. Biol. Chem.* **1995**, *270* (16), 9250–9257.

(103) A novel pathway producing dimethylsulphide in bacteria is widespread in soil environments | Nature Communications https://www.nature.com/articles/ncomms7579 (accessed Jul 5, 2018).

(104) Shimizu, T.; Miyajima, R.; Naruse, N.; Yamaoka, K.; Aihara, K.; Shigenaga, A.; Otaka, A. Facile Preparation of Peptides with C-Terminal N-Alkylamide via Radical-Initiated Dethiocarboxylation. *Chem. Pharm. Bull. (Tokyo)* **2016**, *64* (4), 375–378.

(105) Leman, L. Carbonyl Sulfide-Mediated Prebiotic Formation of Peptides. *Science* **2004**, *306* (5694), 283–286.

(106) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives1Dedicated to Professor Keiji Morokuma in Celebration of His 65th Birthday.1. *J. Mol. Struct. THEOCHEM* **1999**, *461–462*, 1–21.

APPENDIX

Cartesian coordinates for optimized DFT structures of the QDO calculations

1

19

C	0.560185808634	-3.206616654161	0.000000000000
C	1.245040099430	-2.011037850004	0.000000000000
C	0.551358456545	-0.779406182658	0.000000000000
C	-0.864952162732	-0.802513254151	0.000000000000
C	-1.559461863602	-2.029945199197	0.000000000000
C	-0.851607750143	-3.211881643324	0.000000000000
H	1.101108772381	-4.145474810224	0.000000000000
H	2.327891550698	-1.994602730409	0.000000000000
C	1.259128158855	0.490791411404	0.000000000000
H	-2.643739327548	-2.024580115308	0.000000000000
C	-0.928396291695	1.589240390033	0.000000000000
C	0.434198730456	1.666881205112	0.000000000000
N	-1.553277037767	0.384729844737	0.000000000000
O	1.063286571291	2.889249959077	0.000000000000
H	2.017285782038	2.693676237185	0.000000000000
O	2.517491763449	0.608118243300	0.000000000000
H	-2.566152500920	0.363517565274	0.000000000000
H	-1.561695131650	2.465885618973	0.000000000000
H	-1.385665883911	-4.155162819791	0.000000000000

2

18

C	-2.451604810678	0.802172394909	-0.342907537486
C	-1.216123739468	1.438499097080	-0.403732410679
C	-0.024672473928	0.768250059130	-0.091914452661
C	-0.099869450480	-0.598960860127	0.230300835553
C	-1.342078578811	-1.242129092685	0.282042572456
C	-2.513799818877	-0.545311945826	0.012085706927
H	-3.357152906751	1.353339123726	-0.568276958815
H	-1.162846716979	2.487097533129	-0.671885990544
C	1.259291038643	1.589799284679	-0.066436098907
H	-1.376929174298	-2.298043141057	0.527354627717
C	2.238666182422	-1.319259888350	-0.110812580067
N	1.045977090182	-1.390472158004	0.511690927417
O	1.472165746179	2.338492564643	-1.060676318977
H	0.932972166602	-2.141030015167	1.186141690633
H	2.288797710457	-0.592726091153	-0.928960034436

H	-3.467505556813	-1.057771392377	0.062723144755
O	3.204488167172	-2.031572606796	0.181087817808
O	1.999027790971	1.494057521352	0.951971255797

2(H)

19

C	0.392911271709	-2.835054868306	-1.253185875973
C	1.159322436119	-1.739218282017	-0.880157880552
C	0.575277596539	-0.616090451404	-0.271337061169
C	-0.805154476714	-0.641865468412	0.018367026511
C	-1.569809414133	-1.745843250309	-0.368171709170
C	-0.980661627415	-2.828083507750	-1.010533179390
H	0.861260564206	-3.682487396582	-1.738555384539
H	2.219514024711	-1.726693298055	-1.094346837394
C	1.413708287487	0.598660731818	-0.076881191568
H	-2.630164679572	-1.750804168963	-0.142555291750
C	-1.026231293357	1.027925530839	1.815692657996
N	-1.470941762871	0.403214164715	0.702214033947
O	1.012712995242	1.747645558057	-0.137780750213
H	-2.429280986221	0.597026677647	0.427535251226
H	-0.059724711078	0.668912274432	2.187818455738
H	-1.592527975285	-3.673001811735	-1.303970214254
O	-1.655767954122	1.913625964112	2.394679623328
O	2.712111973707	0.307945677212	0.138929464901
H	3.210439370583	1.141362887548	0.190441766187

3

18

C	0.556003660406	-3.209467417848	0.000000000000
C	1.234110919414	-2.011207703668	0.000000000000
C	0.543068946271	-0.771909365397	0.000000000000
C	-0.875927439069	-0.809121256824	0.000000000000
C	-1.566844568235	-2.042517152843	0.000000000000
C	-0.859204822069	-3.222631024555	0.000000000000
H	1.102073946966	-4.145785653255	0.000000000000
H	2.317025545584	-1.995957570791	0.000000000000
C	1.259378083932	0.494097716828	0.000000000000
H	-2.651906485362	-2.037892262249	0.000000000000
C	-0.920353061715	1.582024243193	0.000000000000
C	0.463100471286	1.726187370002	0.000000000000
N	-1.554967936519	0.371428359615	0.000000000000
O	1.047835382181	2.899691513517	0.000000000000
O	2.528900864222	0.543627423740	0.000000000000
H	-2.567087234102	0.353590913140	0.000000000000
H	-1.571579145628	2.446726961183	0.000000000000

H	-1.389044692167	-4.168487812014	0.000000000000

4

20

C	-3.226364425408	0.528684429330	0.362019291910
C	-2.030891911382	1.218190926507	0.380067791816
C	-0.826449293025	0.600932626285	-0.015791646891
C	-0.836271919189	-0.753009967822	-0.441910035535
C	-2.065307877883	-1.445094949387	-0.453460488792
C	-3.230513052779	-0.813910055517	-0.059493693043
H	-4.147188807969	1.009191596840	0.668443569064
H	-1.994702891875	2.252168888186	0.701801992438
C	0.420717659507	1.346346703594	0.004961666408
H	-2.080784085885	-2.480592754095	-0.775183333740
C	1.643734229475	-0.922203722982	-0.617219906688
C	1.685126303551	0.606256943934	-0.464519356060
N	0.299974636887	-1.380980662020	-0.889919348194
O	2.692902778006	1.227704830993	-0.735776421043
O	0.532497708104	2.524072359220	0.344383328113
H	0.217877282907	-2.372242940859	-1.078761409559
H	2.292965222532	-1.235229416559	-1.438073106752
H	-4.161489754909	-1.369854908143	-0.074324396627
O	2.220882615875	-1.440139102599	0.556920853331
O	1.655695583459	-0.588841824905	1.643991649846

4(H)

21

C	3.490365070044	-0.170476257329	0.103089666357
C	2.536315931676	0.819982607268	0.123131679010
C	1.158931550663	0.508859414603	0.023497008639
C	0.755024526241	-0.846807338066	-0.111767218752
C	1.747070183731	-1.850828492425	-0.121182051299
C	3.080884283566	-1.515164333160	-0.019196784034
H	4.542349742735	0.072873037260	0.183594390574
H	2.818620172737	1.861025374950	0.221601454814
C	0.176942740436	1.570376467581	0.046281303895
H	1.442027066092	-2.887073343225	-0.213695888698
C	-1.628371482037	-0.324546316629	0.156860124987
C	-1.302109007277	1.148821713765	-0.068924481132
N	-0.563464056775	-1.207111298429	-0.268831522519
O	-2.161601501317	1.971168941810	-0.280214958897
O	0.419599543710	2.772820579181	0.116695971882
H	-0.758477154950	-2.184065812144	-0.077043562599
H	-1.806659221632	-0.423385047185	1.239363672945
H	3.825361062011	-2.303313277960	-0.030358539836

O	-2.833283789082	-0.598176226310	-0.530563415163
O	-3.343964316710	-1.866683814075	-0.050362892123
H	-4.004524343862	-1.580623579480	0.603659041948

70 (**4(H)** to **6** TS)

21

C	3.438656176816	-0.444554015006	0.172596986243
C	2.580577427250	0.618649774536	0.341208470562
C	1.187635090815	0.452580878819	0.155790587335
C	0.671370061865	-0.833680516369	-0.169763046084
C	1.568776920690	-1.897078133508	-0.379450485859
C	2.924246841377	-1.699996041705	-0.208847227480
H	4.503266860452	-0.316551626616	0.321106102669
H	2.955528960865	1.602109900006	0.594864267338
C	0.363662401794	1.630548540287	0.164446441552
H	1.175296468868	-2.868393227511	-0.654371559198
C	-1.685704865699	-0.198213822622	0.224103059065
C	-1.122728589053	1.467756412332	-0.385083291951
N	-0.675365451737	-1.052025773966	-0.284577450606
O	-1.802719052964	2.208780906847	-0.951866595139
O	0.626306866363	2.763482312910	0.490537130023
H	-0.999953887134	-1.855277619443	-0.815735185526
H	-1.502262947423	0.071951830937	1.279431618477
H	3.602351760444	-2.531185077808	-0.362470489317
O	-2.865996360692	-0.341479508535	-0.188456564949
O	-3.569518062092	-2.226021877386	0.687931322440
H	-4.372284620802	-1.718428316200	0.862683910405

6

19

C	0.452752906032	-2.905526614148	-1.065362675684
C	1.361778527349	-1.921706428453	-0.699838770471
C	0.865700026841	-0.728391084952	-0.164529534682
C	-0.513802250405	-0.555994278695	-0.014070749220
C	-1.434345618038	-1.515234495310	-0.372275501391
C	-0.924353700250	-2.704620378955	-0.904015730949
H	0.808873857298	-3.840338904343	-1.479611061189
H	2.428607699774	-2.060506576359	-0.822120822699
C	1.595659924401	0.443419076108	0.275389686162
H	-2.498579131778	-1.357730965168	-0.248872927164
C	-1.645634398008	0.685608215525	1.885030484464
C	0.540512565694	1.459454241890	0.768604577345
N	-0.856616438740	0.754499081554	0.568969510492
O	0.626756213504	2.539563654331	1.216619372398
O	2.774655067827	0.706227671031	0.305227729653

H	-1.404263662528	1.340373670249	-0.082385968692
H	-1.240391849973	-0.081217888659	2.553146496699
H	-1.614990421144	-3.486017029299	-1.196400479161
O	-2.553578592811	1.420005424090	2.050629351747

71 (**6** to **7** TS)

19

C	-3.129245332880	0.157715246457	-0.054183988959
C	-2.092974135433	1.006145575297	-0.383037544474
C	-0.754688245625	0.567542510030	-0.223535174717
C	-0.475925464818	-0.757183714479	0.196365254239
C	-1.540908465308	-1.579099724533	0.553139794029
C	-2.849096977636	-1.123697137917	0.436031709615
H	-4.153251878259	0.490956715399	-0.163178169357
H	-2.281812431436	2.015128337814	-0.725984608708
C	0.219108511260	1.585676015364	-0.387942984191
H	-1.337185087060	-2.589385947517	0.886241108737
C	1.666352640950	-1.400101745438	-0.851078116001
C	1.698717643456	1.130861083024	0.744910741873
H	1.015969515661	-1.945027813140	0.987501728108
H	1.348635451351	-0.829207600466	-1.733418233696
N	0.840601993219	-1.262537499920	0.250779095704
O	2.588160356624	1.286122646110	1.408337244200
O	0.429012538157	2.620099261595	-0.862605556367
O	2.668667282862	-2.089701594723	-0.831449197196
H	-3.663631902590	-1.785114909750	0.705460662616

7

17

C	0.500508869833	-2.867333622191	-1.074471436270
C	1.273924684806	-1.860621177253	-0.553354016836
C	0.630239486803	-0.715875905642	0.008433265844
C	-0.794031529748	-0.619429043911	0.092349889607
C	-1.532921447070	-1.650521438742	-0.488946691587
C	-0.899178714432	-2.744671297424	-1.059144451542
H	0.970292621849	-3.740618317404	-1.507124354398
H	2.355082696912	-1.898467571719	-0.574551214701
C	1.459895127600	0.327575411271	0.329233688160
H	-2.614340300051	-1.593025074208	-0.462196409629
C	-1.076342008393	1.155056798764	1.793849937894
N	-1.459894827635	0.443042230613	0.684345682702
O	2.179932616669	1.172602047244	0.536839945485
H	-2.425969575567	0.579804639133	0.394280416285
H	-0.119009599590	0.854384287951	2.238045404425
H	-1.506235103747	-3.533114364061	-1.487758298943

O	-1.765321145890	2.038808876653	2.273094910038

72 (**7** to **2(H)** TS)

19

C	-2.783763034375	-0.068644344564	-0.128643798280
C	-1.753528457996	0.842251654697	-0.095358473433
C	-0.423681734751	0.374919436491	0.079535223303
C	-0.129364230039	-1.024317169183	0.148820015898
C	-1.207494398294	-1.906659144640	0.139989662596
C	-2.507969994175	-1.436408048724	0.011582069912
H	-3.802709137780	0.273150747002	-0.254603677658
H	-1.894821521397	1.911659318362	-0.178798762332
C	0.576806352612	1.280651964209	0.385237970804
H	-1.009327098560	-2.969659654029	0.208022265798
C	2.269194341610	-1.115165202470	-0.446565154934
N	1.166109851026	-1.526886129408	0.257364720514
O	1.488052993439	1.841219092929	0.757791044783
H	1.256091323863	-2.433517710004	0.710943781830
H	2.104017606879	-0.264756958760	-1.119090245247
H	-3.322457405604	-2.150719597824	-0.003289967603
O	3.351356030045	-1.671940203960	-0.352337490820
O	-0.678153689821	3.661256903150	-0.546724362170
H	0.195534203317	4.048700046724	-0.418953822962

68 (**4** to **5** TS)

20

C	-2.999126212954	0.939089523415	0.036850110672
C	-1.683565008834	1.399898663689	0.067595130147
C	-0.610967739416	0.517565610595	-0.045780641232
C	-0.861554848963	-0.860349539753	-0.209801946835
C	-2.181297599960	-1.323124933048	-0.236860222919
C	-3.240573981877	-0.427388309027	-0.112863036885
H	-3.824039407964	1.636118243159	0.125809609836
H	-1.475914084067	2.457773332823	0.178986590997
C	0.823772265903	1.031771735691	-0.012576636798
H	-2.369331622128	-2.384656893731	-0.360467337883
C	1.512373180433	-1.313511601327	0.028216106459
C	1.761300602731	0.015640504325	-0.670718739025
N	0.210790934246	-1.757928995567	-0.391775669187
O	2.569000025820	0.238473640163	-1.543711149895
O	1.063898163246	2.277867856333	-0.115200304673
H	0.001935384125	-2.720646899440	-0.153983422063
H	2.258035299374	-2.079226627109	-0.172528705831
H	-4.258206858433	-0.800986616066	-0.138393908228
O	1.586392260455	-0.978344346498	1.431003774473

O	1.402069248264	0.501060651373	1.514838398871

5
20

C	-2.985404402463	0.944428701888	0.009487763261
C	-1.667050995426	1.398389381208	0.043749299647
C	-0.598828008868	0.507969882229	-0.034024616134
C	-0.856692431153	-0.871272821937	-0.172769476851
C	-2.178563390496	-1.327863452172	-0.200980425418
C	-3.233927097088	-0.423662379088	-0.107992930854
H	-3.806929617938	1.648369423093	0.072390530679
H	-1.454304772515	2.457337078700	0.133205584648
C	0.849815833592	1.016328827882	0.028094538048
H	-2.371679792603	-2.390801039419	-0.302832714893
C	1.520097992585	-1.314784558973	0.038150218356
C	1.739119831134	-0.002373242236	-0.699330913207
N	0.212023821485	-1.781002848419	-0.330664352271
O	2.472158833686	0.202397573467	-1.638124656011
O	1.082055768579	2.275647526934	-0.109175900557
H	0.009303204225	-2.728902077211	-0.034002960419
H	2.269350429322	-2.080352354546	-0.148648240125
H	-4.253427346199	-0.792091933068	-0.134195410751
O	1.629021056060	-0.929527666621	1.428295127631
O	1.389522611519	0.549954886102	1.482118950844

69 (**5** to **2** TS)
20
7cisb-ts_final

C	-3.005031783290	0.935227760144	-0.025980380590
C	-1.690885576321	1.399315354226	0.025939103617
C	-0.613489066172	0.515925607753	0.000484104729
C	-0.850543560288	-0.869497929418	-0.113744224207
C	-2.172886996905	-1.333102807957	-0.151896111889
C	-3.238097089986	-0.438535903112	-0.103577755653
H	-3.832789331534	1.634286923968	-0.001339998996
H	-1.488015312989	2.461516561579	0.099253476601
C	0.827904561208	1.038919382985	0.197213850983
H	-2.355321545566	-2.400054145366	-0.232145438244
C	1.562197416107	-1.377918088916	0.038352375094
C	1.738905335684	-0.013799321741	-0.701346331425
N	0.206836593978	-1.789954670582	-0.264086529420
O	2.397722070313	0.281383156550	-1.635445841131
O	1.089983821698	2.284243470133	-0.143482000636
H	-0.010383269348	-2.738274792290	0.020049960269
H	2.291124874264	-2.076726236437	-0.381042798354

H	-4.253227359148	-0.818647378803	-0.136272900292
O	1.866579423947	-1.108028492501	1.329125210271
O	1.292835794348	0.684904549784	1.464979229274

CO
2

O	0.000000000000	0.000000000000	0.036204607822
C	0.000000000000	0.000000000000	1.163795392178

HID (Histidine with hydrogen on the delta nitrogen, side chain)
12

C	-2.175224731015	-0.027413176756	0.000000228604
H	-2.561800173981	-0.545573203023	0.883101009619
H	-2.561800322836	-0.545572888035	-0.883100673754
C	-0.686213164155	0.016463128992	0.000000116772
N	0.102645261747	-1.117548032021	0.000000051722
C	0.208459466559	1.056324107612	0.000000072573
C	1.404820536169	-0.725987483132	-0.000000103275
N	1.513409644505	0.588242294591	-0.000000105033
H	-0.013104418993	2.112409040205	0.000000131193
H	2.219384159463	-1.433290972714	-0.000000205918
H	-2.576001159977	0.987371906740	0.000000448009
H	-0.230907647486	-2.072572682461	0.000000051722

HIE (Histidine with hydrogen on the epsilon nitrogen, side chain)
12

C	-3.580560615403	1.293174144167	1.224278436626
H	-4.248985409481	0.426864518662	1.192102253954
H	-2.904737922445	1.150667098367	2.073557431366
C	-4.349349576920	2.567084555488	1.350641019033
N	-5.189914339947	2.819157342893	2.425540778909
C	-4.366417262846	3.639564930543	0.496007766408
C	-5.696163891888	4.020876375344	2.210801684841
N	-5.230058931537	4.558559229203	1.058462098160
H	-3.857701864153	3.828414620766	-0.434872467764
H	-6.394153459462	4.539500656813	2.849392224722
H	-2.981359273500	1.291859375750	0.311556735292
H	-5.471976147062	5.465558869504	0.682616045408

HIP (Histidine with hydrogens on both nitrogens, side chain)
13

C	-3.572127103659	1.275153920453	1.226214246798
H	-4.254122212628	0.421128929454	1.197972763996
H	-2.904786578661	1.147671138892	2.082764513483
C	-4.323318857697	2.553178076487	1.322362642933

N	-5.175855576959	2.850919517443	2.378019469246
C	-4.367793896980	3.638667695785	0.498844259740
H	-5.367321813506	2.244173373662	3.168238071476
C	-5.715385782102	4.058097327178	2.209552367519
N	-5.235491063572	4.551398009710	1.072626291241
H	-3.864851129836	3.837060180318	-0.431068470660
H	-6.411152034777	4.543154467355	2.872315906690
H	-2.972852778544	1.271009008332	0.315705363486
H	-5.478241322311	5.460486756760	0.696334612688

O_2
2

O	0.000000000000	0.000000000000	-0.002174507798
O	0.000000000000	0.000000000000	1.202174507798

OH^-
2

O	0.000000000000	0.000000000000	-0.131741607551
H	0.000000000000	0.000000000000	0.831741607551

Asp (Side chain)
7

C	-2.543572453279	-0.453179512176	-1.326947032394
H	-3.484901288246	-0.506584781168	-1.874427900334
H	-2.054393193854	-1.430826051939	-1.357270185067
C	-2.750964761390	-0.032114695410	0.124854216502
O	-3.932745210843	0.183687243858	0.527713671920
O	-1.714833503993	0.076174909686	0.848910771761
H	-1.877856588395	0.257782887149	-1.823957542388

Aspn (protonated Asp, side chain)
8

H	-0.714666434155	-0.145224933114	0.410940502125
C	-2.329512342147	-0.359485915347	-1.366255432494
H	-3.259669121468	-0.402667495677	-1.928355559146
H	-1.791538834249	-1.306354830224	-1.464857453902
C	-2.613591831701	-0.089783356790	0.080115721475
O	-3.719522775815	0.060389156919	0.561405260766
O	-1.539221251743	-0.010230704684	0.902776452802
H	-1.688782408721	0.430122078917	-1.768380491626

1 with key active site residues
130

N	7.848471357862	74.393251629358	21.490129366540
C	8.106013627541	73.802086435294	22.688005277677

C	7.131400134956	73.752524691538	23.665621913595
C	5.827635716524	74.283539745667	23.418301462548
C	4.393147439730	75.595512806786	21.847442662493
C	4.204382231521	76.213287931304	20.630638417535
C	5.236783317865	76.210858403138	19.669549303498
C	6.445566725048	75.605966552983	19.939750499457
C	6.650901991948	74.982840666947	21.189374820480
C	5.617467464944	74.960080753587	22.152524592237
C	9.470322947165	73.231399136654	22.908153839352
O	7.430331356393	73.161614177304	24.832992511217
O	4.905216151892	74.152797518965	24.277293205175
H	6.798270142335	73.353825012447	25.645057352495
H	3.611913989258	75.588379960727	22.596817922420
H	3.261120835689	76.697786649191	20.408182340358
H	5.082646298297	76.696798681367	18.712389062423
H	7.250187554881	75.614584993857	19.212552439475
H	9.407572902860	72.172635924495	23.161008706107
H	10.100733381303	73.340084839494	22.026961484904
H	9.949838639921	73.730281721527	23.754539761525
H	8.529191381712	74.246574613639	20.742615760908
N	6.362888078958	73.475218442826	27.137531591940
C	7.396993534627	74.869000788665	30.407993342824
C	6.809327507182	74.172564871011	29.230311792668
N	5.655585055170	73.412506101035	29.239112173223
C	7.229907500396	74.196553808098	27.925173233536
C	5.434210183678	73.020194281985	27.967362241920
H	4.603281105778	72.398702595207	27.684826094615
H	8.069052222056	74.726605384916	27.503480359229
H	5.054827488896	73.061639633177	30.056572575081
H	6.736899468815	75.658621730952	30.779320580393
H	8.344998030626	75.324109128359	30.125623853218
H	7.574097595517	74.181124559279	31.240743720266
C	4.959005432275	70.081002097741	30.856999386834
C	4.588986275743	71.287990968880	31.707994556642
O	4.099007896101	72.312009724092	31.133989339804
O	4.720009171405	71.172003600370	32.931999359383
H	5.645140154010	70.351859339600	30.041033477312
H	4.055711591725	69.664322181366	30.387786416819
H	5.425526700475	69.310291571575	31.474311630451
C	9.457957316118	72.840156903990	18.471896942161
O	9.560988046995	73.531864167965	19.481109020946
N	10.345997575905	72.843003339212	17.485997825187
C	13.728999058117	72.340001097411	21.093000036043
C	12.756004988522	71.480994265835	21.820999424831
N	12.739002113744	71.401998277814	23.193001756355

C	11.781007049865	70.655993030126	21.376001181802
C	11.762992281285	70.595013227799	23.568995371120
N	11.163969233084	70.131008449556	22.485009808405
C	7.376000267865	66.810995968153	25.829002888193
C	6.417003988847	67.754003890323	25.125003587674
O	6.454999400804	67.903999309190	23.890999897583
C	8.805008782867	67.393002113615	25.812995339570
C	8.819957471146	68.832999370272	26.209999095825
N	8.634990613022	69.237999579831	27.524001770717
C	8.828078479964	69.955988687211	25.454001158456
C	8.595008649300	70.556994389925	27.553995813128
N	8.701980346887	71.012009676120	26.314994760560
N	5.552992963394	68.376001712672	25.932013626646
C	4.597004758624	69.330996288143	25.478971531635
C	3.861002349327	68.837994312034	24.250997721966
O	3.293000858638	67.716999555040	24.287999700360
C	5.265983338310	70.730990635844	25.239995854215
O	4.229029470835	71.702014187294	25.039056977690
N	3.875010192018	69.639007378498	23.171994867720
C	3.123978122090	69.342987917314	21.957003108351
C	2.765022089562	70.606011593436	21.160004292496
C	1.940994903165	71.556990218038	21.957003115867
N	0.578998480025	71.393999138143	22.161000530586
C	2.311995182877	72.632002456616	22.692992683663
C	0.154003621207	72.325998811469	22.996998212638
N	1.183000103384	73.107003769382	23.320004175939
H	8.456981427652	71.173464840484	28.425398145663
C	8.537000012562	78.069999925808	29.325000011759
C	7.616000591098	77.736996856420	28.206000329379
C	6.289983572443	77.498003615133	28.287006008500
C	7.962004607175	77.604974152558	26.826003918774
N	5.777000376730	77.245082917064	27.046944116131
C	6.791020791773	77.295001232415	26.132022006415
C	9.158999896550	77.700000650516	26.118999853536
C	6.771000294633	77.063993164927	24.753000604939
C	9.137999210185	77.508004119872	24.733999407678
C	7.960001144415	77.186995631974	24.073001192197
O	2.043948875592	72.515229919894	26.835780672175
O	2.920968487426	77.013924000241	26.802981075743
O	2.504115850703	74.728676902793	25.336270808366
H	11.488976240665	70.345641574006	24.579166532132
H	11.447084575259	70.407123499362	20.381478740752
H	5.845151107847	76.815920640522	24.248996228489
H	7.977953136907	77.030524300970	22.997658741931
H	10.064596342431	77.590273495580	24.172047985989

H	10.091375404558	77.934937705560	26.625155454868
H	4.775285575033	77.090485786487	26.871377977427
H	5.645929193360	77.500980343297	29.153119947070
H	8.840037806188	70.106204189225	24.388483610097
H	10.243927253056	72.255454722538	16.672996316072
H	8.606184405531	72.162925359587	18.311001087363
H	11.163266688179	73.434810280266	17.552628404604
H	14.759292433550	72.077236416779	21.359405892530
H	13.589869246841	73.395944394131	21.353538200730
H	13.618297338464	72.237075610125	20.007724315040
H	3.417416842721	74.592842484347	24.972774023593
H	2.284471984832	74.006223901298	25.996615137454
H	2.667726474551	76.190351498992	26.287364905609
H	2.404863705455	77.775703556473	26.488108423035
H	2.675204836423	71.970930443261	26.300017994133
H	2.162065529475	72.368038389160	27.791441626834
H	9.443197621651	66.808573534583	26.485178501952
H	9.210228648484	67.293052941497	24.798386126159
H	7.069127288911	66.641217054609	26.867075418892
H	7.354074827704	65.846985089608	25.309221636319
H	5.630020018646	68.225326400790	26.926396425129
H	4.591530644077	72.625655661997	24.884283266811
H	5.939148063787	70.670760425107	24.380636830416
H	5.868118515743	70.975624716730	26.119838020536
H	3.846433549365	69.457659806199	26.267755482741
H	4.092703335613	70.614524520213	23.332582568434
H	2.205013942644	68.812902254836	22.234370953356
H	3.707434376346	68.656365473202	21.326918756075
H	3.684311222649	71.102798157418	20.817974577386
H	2.212208620998	70.292884822905	20.266855647846
H	3.267633533774	73.085406551851	22.867814060794
H	-0.854013426995	72.451531107381	23.352324788555
H	10.399677137516	69.472622806602	22.459467014768
H	1.249254279192	73.850185193083	24.016099763462
H	8.545851322792	71.956982401747	25.979363603572
H	9.006234985881	79.056891249713	29.194888310415
H	9.359757252988	77.346436507424	29.420906404650
H	8.008612580782	78.086541672398	30.284560378256

1 to **3** TS with key active site residues

130

N	-3.411022525343	1.634430582356	0.805959465050
C	-2.524776002721	1.210182043934	-0.137126812750
C	-1.188988181485	1.009697461594	0.181427481847
C	-0.747994230785	1.217760194744	1.529137640989

C	-1.339090404508	2.050707600944	3.810891188274
C	-2.272309845186	2.517897458215	4.709332982536
C	-3.618220411769	2.674917637207	4.314033646432
C	-4.009024296390	2.378739669236	3.026268568635
C	-3.054880744032	1.913216234939	2.095298303014
C	-1.707899701721	1.729198592360	2.483983649894
C	-3.039838286038	0.989515563299	-1.522373186620
O	-0.345146744676	0.599403596345	-0.758013032662
O	0.451866393587	0.949526154477	1.854521029330
H	0.821848906212	0.711226938999	-0.632473955570
H	-0.303060732572	1.925282096001	4.098802298980
H	-1.975936401423	2.760844597525	5.722911318083
H	-4.349797971673	3.040696761633	5.026180126131
H	-5.037455522031	2.511930161611	2.708678859381
H	-2.837295310266	-0.030592939023	-1.848255443189
H	-4.112429961163	1.168403706335	-1.590691766669
H	-2.523824927699	1.652145513781	-2.222698057313
H	-4.403855011695	1.603807903893	0.567996308776
N	2.094471738539	0.859864344198	-0.930329229581
C	4.505105000000	2.634373000000	-3.081701000000
C	3.732952003765	1.750484294084	-2.167779942195
N	4.253155614851	0.738337516792	-1.378767145154
C	2.396783226206	1.806966842399	-1.877898091061
C	3.234510332218	0.235504265060	-0.656317490648
H	3.342542086177	-0.574575806544	0.041441381484
H	1.650905390220	2.490997208979	-2.247295787522
H	5.239098199610	0.301035908682	-1.363113713775
H	5.209304418044	3.264417692757	-2.530642910217
H	3.818840494645	3.283679368916	-3.622318483663
H	5.088602716450	2.059195410390	-3.807173223473
C	5.815677000000	-2.550560000000	-2.397042000000
C	6.809753000000	-1.410044000000	-2.227283000000
O	6.619951000000	-0.569364000000	-1.291655000000
O	7.796068000000	-1.410669000000	-2.973000000000
H	4.784656000000	-2.181385000000	-2.500132000000
H	5.834438000000	-3.197240000000	-1.507547000000
H	6.070112000000	-3.148018000000	-3.275117000000
C	-6.969244000000	0.357305000000	0.658439000000
O	-6.105039000000	1.119147000000	0.233316000000
N	-8.264137000000	0.502495000000	0.408339000000
C	-6.860556000000	1.045459000000	-4.329148000000
C	-5.798613000000	0.032826000000	-4.081267000000
N	-4.608313000000	0.043208000000	-4.768288000000
C	-5.748129000000	-1.025079000000	-3.240102000000
C	-3.846143000000	-0.943015000000	-4.330506000000

N	-4.513589000000	-1.605153000000	-3.400322000000
C	0.063086000000	-5.490167000000	-2.683249000000
C	-0.010443000000	-4.846592000000	-1.310312000000
O	-1.087510000000	-4.774043000000	-0.692760000000
C	-0.627155000000	-4.592282000000	-3.731306000000
C	-0.203293000000	-3.165707000000	-3.602268000000
N	1.050080000000	-2.729018000000	-4.006248000000
C	-0.792627000000	-2.124678000000	-2.968511000000
C	1.176491000000	-1.457005000000	-3.677239000000
N	0.079364000000	-1.072785000000	-3.042972000000
N	1.155984000000	-4.391956000000	-0.844033000000
C	1.298742000000	-3.719408000000	0.404356000000
C	0.573233000000	-4.451277000000	1.513713000000
O	0.819759000000	-5.669076000000	1.705584000000
C	0.844301000000	-2.220734000000	0.294335000000
O	1.246340000000	-1.534906000000	1.488790000000
N	-0.318066000000	-3.742597000000	2.227453000000
C	-1.012830000000	-4.286056000000	3.389278000000
C	-1.446194000000	-3.197821000000	4.382870000000
C	-0.288256000000	-2.415391000000	4.898204000000
N	0.556782000000	-2.878696000000	5.895438000000
C	0.229122000000	-1.237035000000	4.475391000000
C	1.548492000000	-2.018706000000	6.053757000000
N	1.362906000000	-0.998929000000	5.217845000000
H	2.036585000000	-0.833563000000	-3.850409000000
C	3.195953000000	5.934232000000	-2.741039000000
C	2.666633000000	5.321489000000	-1.493779000000
C	3.382728000000	4.784729000000	-0.483286000000
C	1.292622000000	5.182890000000	-1.127141000000
N	2.550315000000	4.337120000000	0.502417000000
C	1.256853000000	4.562169000000	0.122788000000
C	0.090585000000	5.508301000000	-1.752358000000
C	0.059995000000	4.241964000000	0.771222000000
C	-1.108573000000	5.225310000000	-1.090946000000
C	-1.112995000000	4.594431000000	0.145831000000
O	3.938656000000	-1.138301000000	2.610509000000
O	3.748081000000	3.428351000000	2.953818000000
O	2.547822000000	1.016736000000	3.493273000000
H	-2.851152000000	-1.182033000000	-4.663043000000
H	-6.457134000000	-1.410571000000	-2.525103000000
H	0.070191000000	3.751250000000	1.736547000000
H	-2.061564000000	4.375549000000	0.629021000000
H	-2.051219000000	5.484687000000	-1.565585000000
H	0.078196000000	5.987944000000	-2.727295000000
H	2.888085000000	3.941100000000	1.389681000000

H	4.452888000000	4.694329000000	-0.375797000000
H	-1.710984000000	-2.046835000000	-2.412998000000
H	-8.952330000000	-0.145243000000	0.759664000000
H	-6.725328000000	-0.509646000000	1.290224000000
H	-8.577528000000	1.272171000000	-0.167847000000
H	-7.159483000000	1.049364000000	-5.383733000000
H	-6.501415000000	2.054408000000	-4.094001000000
H	-7.750351000000	0.847748000000	-3.720998000000
H	1.769543000000	1.075138000000	2.880118000000
H	3.184083000000	0.308558000000	3.178257000000
H	3.377967000000	2.536076000000	3.227908000000
H	3.779393000000	4.025591000000	3.720721000000
H	3.127537000000	-1.554055000000	2.221892000000
H	4.697378000000	-1.190055000000	2.001767000000
H	-0.399418000000	-4.965233000000	-4.736152000000
H	-1.712683000000	-4.661683000000	-3.588347000000
H	1.103439000000	-5.657822000000	-2.983496000000
H	-0.434424000000	-6.464826000000	-2.633308000000
H	1.968455000000	-4.454137000000	-1.438556000000
H	0.988315000000	-0.564595000000	1.480371000000
H	-0.238058000000	-2.178626000000	0.145726000000
H	1.320193000000	-1.784205000000	-0.588505000000
H	2.362640000000	-3.719891000000	0.668370000000
H	-0.228054000000	-2.734783000000	2.195488000000
H	-0.347472000000	-4.996886000000	3.893697000000
H	-1.890405000000	-4.857300000000	3.053987000000
H	-2.169674000000	-2.523502000000	3.902510000000
H	-1.962392000000	-3.690143000000	5.215050000000
H	-0.068089000000	-0.562014000000	3.697319000000
H	2.365680000000	-2.109278000000	6.748100000000
H	-4.195114000000	-2.424320000000	-2.904535000000
H	1.977340000000	-0.216401000000	4.991179000000
H	-0.075433000000	-0.214988000000	-2.523146000000
H	2.909912000000	6.992667000000	-2.835701000000
H	2.824815000000	5.431303000000	-3.646020000000
H	4.290138000000	5.890256000000	-2.773417000000

3 with key active site residues

130

N	-3.396790083277	1.614350411234	0.849828881086
C	-2.537852257242	1.162567810866	-0.106708444555
C	-1.183665773365	0.948762462044	0.171256020946
C	-0.722951232091	1.196162415585	1.516163234256
C	-1.256838076279	2.067350666575	3.799139286959
C	-2.167010576635	2.552437113553	4.709836750439

C	-3.523531262793	2.708184162427	4.344194163131
C	-3.942764772225	2.392472706204	3.070850892186
C	-3.012563641181	1.907701546971	2.124623461645
C	-1.653958369336	1.719725666844	2.483386076578
C	-3.085795010853	0.926691547628	-1.475289176710
O	-0.357959243412	0.535153910202	-0.745738657982
O	0.498964199844	0.949868489433	1.813353352297
H	1.200623888255	0.646510580485	-0.562739636070
H	-0.215442322418	1.944538995027	4.068488921858
H	-1.846353295497	2.810998696299	5.712343935480
H	-4.237297548937	3.087758439755	5.067169475670
H	-4.978038893671	2.525296327766	2.774536939999
H	-2.873261250992	-0.092616536721	-1.797773003648
H	-4.162187256737	1.089607818992	-1.519502200691
H	-2.595998368905	1.588632974086	-2.195904789493
H	-4.392194925239	1.606015697531	0.625193723661
N	2.203684255643	0.824794803694	-0.893688525055
C	4.505104663974	2.634371290703	-3.081701264763
C	3.780413453495	1.732639868224	-2.150188291940
N	4.330205732378	0.708135267342	-1.389333844066
C	2.454151125218	1.790343338963	-1.831514847294
C	3.346326686048	0.187368892842	-0.648059995711
H	3.457479770222	-0.634917640627	0.032754932584
H	1.685378489558	2.467888477833	-2.159728121711
H	5.342352439237	0.244962803797	-1.381880108526
H	5.210238701274	3.277339663825	-2.547988630860
H	3.789577682023	3.269320271072	-3.600128399074
H	5.078987305369	2.072347134384	-3.824644304945
C	5.815676466227	-2.550559444330	-2.397042607608
C	6.809745770899	-1.410047939386	-2.227280928198
O	6.619961497629	-0.569358026871	-1.291658098796
O	7.796069303839	-1.410670626320	-2.972998274139
H	4.784656000000	-2.181385000000	-2.500132000000
H	5.834438000000	-3.197240000000	-1.507547000000
H	6.070112000000	-3.148018000000	-3.275117000000
C	-6.969246927979	0.357341009444	0.658497583847
O	-6.105044700175	1.119124961491	0.233277697795
N	-8.264136323536	0.502490972901	0.408333166671
C	-6.860556000000	1.045459000000	-4.329148000000
C	-5.798612055080	0.032827504069	-4.081265165209
N	-4.608309550604	0.043213503522	-4.768281607622
C	-5.748142451068	-1.025103444678	-3.240117161459
C	-3.846154169000	-0.943037712318	-4.330530178339
N	-4.513571059489	-1.605121847181	-3.400313754697
C	0.063085884515	-5.490166890272	-2.683249254685

C	-0.010443000000	-4.846592000000	-1.310312000000
O	-1.087510000000	-4.774043000000	-0.692760000000
C	-0.627155000000	-4.592282000000	-3.731306000000
C	-0.203293000000	-3.165707000000	-3.602268000000
N	1.050080000000	-2.729018000000	-4.006248000000
C	-0.792627000000	-2.124678000000	-2.968511000000
C	1.176491000000	-1.457005000000	-3.677239000000
N	0.079364000000	-1.072785000000	-3.042972000000
N	1.155984000000	-4.391956000000	-0.844033000000
C	1.298742000000	-3.719408000000	0.404356000000
C	0.573233000000	-4.451277000000	1.513713000000
O	0.819759000000	-5.669076000000	1.705584000000
C	0.844294091004	-2.220736514396	0.294338768603
O	1.246356086305	-1.534900740809	1.488778545404
N	-0.318066000000	-3.742597000000	2.227453000000
C	-1.012830000000	-4.286056000000	3.389278000000
C	-1.446194000000	-3.197821000000	4.382870000000
C	-0.288256000000	-2.415391000000	4.898204000000
N	0.556782000000	-2.878696000000	5.895438000000
C	0.229122000000	-1.237035000000	4.475391000000
C	1.548492000000	-2.018706000000	6.053757000000
N	1.362906000000	-0.998929000000	5.217845000000
H	2.036585000000	-0.833563000000	-3.850409000000
C	3.195953000000	5.934232000000	-2.741039000000
C	2.666633000000	5.321489000000	-1.493779000000
C	3.382728000000	4.784729000000	-0.483286000000
C	1.292622000000	5.182890000000	-1.127141000000
N	2.550315000000	4.337120000000	0.502417000000
C	1.256853000000	4.562169000000	0.122788000000
C	0.090585000000	5.508301000000	-1.752358000000
C	0.059995000000	4.241964000000	0.771222000000
C	-1.108573000000	5.225310000000	-1.090946000000
C	-1.112995000000	4.594431000000	0.145831000000
O	3.938656000000	-1.138301000000	2.610509000000
O	3.748081000000	3.428351000000	2.953818000000
O	2.547826837398	1.016741642967	3.493267685648
H	-2.851147202961	-1.182025370098	-4.663034128177
H	-6.457127898838	-1.410561149908	-2.525091641050
H	0.070191000000	3.751250000000	1.736547000000
H	-2.061564000000	4.375549000000	0.629021000000
H	-2.051219000000	5.484687000000	-1.565585000000
H	0.078196000000	5.987944000000	-2.727295000000
H	2.888085000000	3.941100000000	1.389681000000
H	4.452888000000	4.694329000000	-0.375797000000
H	-1.710983386855	-2.046836117834	-2.413000240644

H	-8.952330000000	-0.145243000000	0.759664000000
H	-6.725325757102	-0.509658928711	1.290205391915
H	-8.577528000000	1.272171000000	-0.167847000000
H	-7.159483000000	1.049364000000	-5.383733000000
H	-6.501415000000	2.054408000000	-4.094001000000
H	-7.750351000000	0.847748000000	-3.720998000000
H	1.769539135376	1.075127561064	2.880128481009
H	3.184085936030	0.308562299345	3.178253265187
H	3.377964447576	2.536077915016	3.227910979105
H	3.779393000000	4.025591000000	3.720721000000
H	3.127537000000	-1.554055000000	2.221892000000
H	4.697378000000	-1.190055000000	2.001767000000
H	-0.399418000000	-4.965233000000	-4.736152000000
H	-1.712683000000	-4.661683000000	-3.588347000000
H	1.103439000000	-5.657822000000	-2.983496000000
H	-0.434424000000	-6.464826000000	-2.633308000000
H	1.968455000000	-4.454137000000	-1.438556000000
H	0.988301929536	-0.564603307610	1.480380145497
H	-0.238058000000	-2.178626000000	0.145726000000
H	1.320193000000	-1.784205000000	-0.588505000000
H	2.362640000000	-3.719891000000	0.668370000000
H	-0.228054000000	-2.734783000000	2.195488000000
H	-0.347472000000	-4.996886000000	3.893697000000
H	-1.890405000000	-4.857300000000	3.053987000000
H	-2.169674000000	-2.523502000000	3.902510000000
H	-1.962392000000	-3.690143000000	5.215050000000
H	-0.068089000000	-0.562014000000	3.697319000000
H	2.365680000000	-2.109278000000	6.748100000000
H	-4.195118567622	-2.424319599799	-2.904531405296
H	1.977340000000	-0.216401000000	4.991179000000
H	-0.075433000000	-0.214988000000	-2.523146000000
H	2.909912000000	6.992667000000	-2.835701000000
H	2.824815000000	5.431303000000	-3.646020000000
H	4.290138000000	5.890256000000	-2.773417000000